SISTEMAS DE COMUNICACIÓN

SISTEMAS DE COMUNICACIÓN

Simon Haykin
McMaster University

LIMUSA WILEY

LC	DEWEY	CUTTER
TK5101	621.38415	H 39s

VERSIÓN AUTORIZADA EN ESPAÑOL DE LA OBRA ORIGINALMENTE PUBLICADA EN INGLÉS POR JOHN WILEY & SONS, CON EL TÍTULO:
COMMUNICATION SYSTEMS

© 2001, JOHN WILEY & SONS, INC. NEW YORK, CHICHESTER, BRISBANE, SINGAPORE, TORONTO Y WEINHEIM

CON LA COLABORACIÓN EN LA TRADUCCIÓN DE:
GABRIEL NAGORE CÁZARES

REVISIÓN:
FRANCISCO RODRÍGUEZ RAMÍREZ
INGENIERO MECÁNICO ELECTRICISTA DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO. POSGRADO EN MICROPROCESADORES Y ELECTRÓNICA DIGITAL POR LA UNIVERSITY OF MANCHESTER, INSTITUTE OF SCIENCE & TECHNOLOGY. PROFESOR DEL DEPARTAMENTO DE INGENIERÍA DE CONTROL DE LA FACULTAD DE INGENIERÍA DE LA UNAM.

LA PRESENTACIÓN Y DISPOSICIÓN EN CONJUNTO DE

SISTEMAS DE COMUNICACIÓN

SON PROPIEDAD DEL EDITOR. NINGUNA PARTE DE ESTA OBRA PUEDE SER REPRODUCIDA O TRANSMITIDA, MEDIANTE NINGÚN SISTEMA O MÉTODO, ELECTRÓNICO O MECÁNICO (INCLUYENDO EL FOTOCOPIADO, LA GRABACIÓN O CUALQUIER SISTEMA DE RECUPERACIÓN Y ALMACENAMIENTO DE INFORMACIÓN), SIN CONSENTIMIENTO POR ESCRITO DEL EDITOR.

DERECHOS RESERVADOS:

© 2002, EDITORIAL LIMUSA, S.A. DE C.V.

GRUPO NORIEGA EDITORES
BALDERAS 95, MÉXICO, D.F.
C.P. 06040
(5) 8503-80-50
01(800) 7-06-91-00
(5) 512-29-03
limusa@noriega.com.mx
www.noriega.com.mx

CANIEM NÚM. 121

PRIMERA EDICIÓN
HECHO EN MÉXICO
ISBN 968-18-6307-0

A la memoria adorable de Vera

PREFACIO

El plan de estudios en ingeniería eléctrica ha experimentado cambios radicales durante las últimas dos décadas y continúa por ese camino. Un programa de estudios moderno en esta especialidad incluye los siguientes cursos introductorios:

- *Señales y sistemas*. Proporciona un tratamiento balanceado e integrado de señales y sistemas en tiempo continuo y en tiempo discreto. La transformada de Fourier (en sus diferentes formas), la transformada de Laplace y la transformada z se abordan en detalle. Por lo general, el curso incluye también un tratamiento elemental de los sistemas de comunicación.
- *Probabilidad y procesos aleatorios*. Desarrolla una comprensión intuitiva de las variables aleatorias continuas y discretas e introduce después la noción de proceso aleatorio y sus características.

En general, estos dos cursos introductorios conducen a un curso avanzado de sistemas de comunicación, por lo que este libro se ha escrito considerando dicho antecedente y objetivo fundamental. En términos simples, ofrece un tratamiento moderno de los sistemas de comunicación a un nivel adecuado para un curso universitario de uno o dos semestres. Se subrayan los fundamentos estadísticos de la teoría de comunicaciones con aplicaciones.

El material se presenta en una secuencia lógica y se ilustra con ejemplos; el objetivo general es ayudar a que el estudiante desarrolle una comprensión intuitiva de la teoría que se explica. Salvo el capítulo Antecedentes y preliminares, todos los demás finalizan con numerosos problemas ideados no sólo para auxiliar a los estudiantes a evaluar su comprensión del material que se abordó en el capítulo, sino también para motivarlos a fin de que amplíen el estudio de dicho material. En cada capítulo se incluye una sección de notas y bibliografía con lecturas recomendadas. Las secciones o subsecciones que es posible omitir sin pérdida de continuidad se identifican con un pie de página.

Un rasgo distintivo de este libro es la inclusión de ocho experimentos de computadora en los que se utiliza MATLAB. Este conjunto de experimentos ofrece las bases de un "Laboratorio de software"; cada experimento ha sido ideado para ampliar el material que se presenta en el capítulo correspondiente. Más importante es el hecho de que los experimentos aprovechan las capacidades únicas de MATLAB de una manera instructiva. Los códigos de MATLAB correspondientes a estos experimentos pueden obtenerse en el sitio: **http://www.wiley.com/college/haykin/.**

El capítulo Antecedentes y preliminares presenta material introductorio y motivacional, y prepara el camino para una exposición detallada de los diversos aspectos de los sistemas de comunicación en los 10 capítulos subsecuentes. El material en éstos se organiza del modo siguiente:

- El capítulo 1 presenta un tratamiento detallado de los *procesos aleatorios* o *estocásticos*, destacando en particular su caracterización parcial (es decir, estadística de segundo orden). En efecto, la exposición se restringe a los procesos estacionarios en sentido amplio. Se describen en detalle las propiedades de correlación y los espectros de potencia de un

proceso aleatorio. Los procesos gaussianos y el ruido de banda angosta sobresalen en el estudio de los sistemas de comunicación, de ahí su estudio en la última parte del capítulo. Esto último conduce de manera natural a las distribuciones de Rayleigh y Rician que surgen en un ambiente de comunicaciones.

- El capítulo 2 presenta un estudio integrado de la *modulación de onda continua* (es decir, *comunicaciones analógicas*) y sus diferentes tipos, como se describe a continuación:
  i) *Modulación de amplitud*: puede asumir una de las siguientes formas (dependiendo de cómo se especifiquen las características espectrales de la onda modulada):
    - Modulación de amplitud total
    - Modulación de doble banda lateral-portadora suprimida
    - Modulación en amplitud por cuadratura
    - Modulación de banda lateral única
    - Modulación de banda lateral residual

  ii) *Modulación angular*: asume una de dos formas interrelacionadas:
    - Modulación en fase
    - Modulación en frecuencia

  Se explican las características en el dominio del tiempo y espectrales de estas ondas moduladas y los métodos para su generación y detección, así como los efectos del ruido del canal en sus desempeños.

- El capítulo 3 abarca la *modulación por pulsos* y explica el procesamiento del muestreo, la cuantificación y la codificación que son fundamentales para la transmisión digital de señales analógicas. Este capítulo puede considerarse como la transición de las comunicaciones analógicas a las digitales. De manera específica, se presentan los siguientes tipos de modulación por pulsos:
  i) *Modulación por pulsos analógicos*, donde únicamente el tiempo se representa en forma discreta; ésta incluye las siguientes formas especiales:
    - Modulación por amplitud de pulsos
    - Modulación por ancho (duración) de pulso
    - Modulación por posición de pulso

  Las características de la modulación por amplitud de pulsos se explican en detalle, ya que ésta constituye la base para todas las formas de modulación por pulsos, sean de tipo analógico o digital.
  ii) *Modulación por pulsos digitales*, en la cual tanto el tiempo como la amplitud de la señal se representan en forma discreta; comprende las siguientes formas especiales:
    - Modulación por pulsos codificados
    - Modulación delta
    - Modulación diferencial por codificación de pulsos

  En la modulación delta, la velocidad de muestreo aumenta bastante con respecto a la que se usa en la modulación por pulsos codificados para simplificar la implementación del sistema. En contraste, en la modulación diferencial por codificación de pulsos, la velocidad de muestreo se reduce utilizando un predictor que explota las propiedades de correlación de la señal que contiene la información.
  iii) El estándar de codificación de audio MPEG, que incluye un modelo psicoacústico como elemento clave en el diseño de un codificador.

- El capítulo 4 aborda la *transmisión de pulsos en banda base*, la cual tiene que ver con la transmisión de señales moduladas por amplitud de base en su forma de banda base. Se analizan dos importantes aspectos: el efecto del ruido del canal y el ancho de banda limitado del canal en el desempeño de un sistema de comunicación digital. Suponiendo que el ruido

del canal es aditivo y blanco, este efecto se minimiza empleando un *filtro de acoplamiento*, el cual resulta fundamental en el diseño de los receptores de comunicación. Con respecto al ancho de banda limitado del canal, éste se manifiesta en la forma de un fenómeno conocido como *interferencia de intersímbolo*. Para combatir los efectos de degradación de esta interferencia dependiente de la señal, es posible utilizar un *filtro de formación de pulso* o un *codificador/decodificador relativo*; se discuten ambos métodos. El capítulo incluye una explicación de las líneas de suscriptor digitales para la comunicación directa entre un suscriptor y un proveedor de servicios de Internet. A lo anterior sigue la deducción del receptor lineal óptimo para combatir los efectos combinados del ruido del canal y la interferencia de intersímbolo, la cual conduce a un estudio introductorio de la *ecualización adaptativa*.

- El capítulo 5 presenta el *análisis espacial de señales* en un canal con ruido blanco gaussiano aditivo. En particular, se establecen los fundamentos para la representación geométrica de señales con energía finita. Se obtiene el receptor de correlación, y se demuestra su equivalencia con el filtro de acoplamiento. Concluye con una explicación de la probabilidad de error y su cálculo aproximado.
- En el capítulo 6 se estudia la *transmisión de datos pasobanda*, donde se recurre a una onda portadora senoidal para facilitar la transmisión de una onda modulada digitalmente por un canal pasobanda. Este capítulo parte de la interpretación geométrica de señales presentada en el capítulo 5. En particular, se evalúa el efecto del ruido del canal en el desempeño de sistemas de comunicación digitales, utilizando las siguientes técnicas de modulación:
  i) *Corrimiento de fase por llaveo*, que es la contraparte digital de la modulación en fase, en la cual la fase de la portadora asume uno de los valores discretos de un conjunto preestablecido.
  ii) Esquemas híbridos de modulación en amplitud/fase que incluyen la *modulación en amplitud por cuadratura* (QAM) y la *modulación en amplitud/fase sin portadora* (CAP).
  iii) *Corrimiento de frecuencia por llaveo*, que es la contraparte digital de la modulación en frecuencia, en la cual la frecuencia de la portadora asume un valor discreto de un conjunto preestablecido.
  iv) Modulación multicanal genérica, seguida por *multitonos discretos*, cuyo uso se ha estandarizado en las líneas de suscriptor digitales asimétricas.

  En un sistema de comunicación digital, la temporización lo es todo, lo cual significa que el receptor debe *estar sincronizado* con el transmisor. En este contexto, se dice que el receptor es coherente o no coherente. En un *receptor coherente*, se toman provisiones para recuperar tanto la fase de la portadora como la temporización del símbolo. En un *receptor no coherente*, se ignora la fase de la portadora y se toman provisiones sólo para la temporización del símbolo. Una estrategia de este tipo depende del hecho de que la fase de la portadora puede ser aleatoria, lo que implica que la recuperación de esta última constituya una propuesta costosa. En la última parte del capítulo se explican las técnicas de sincronización, destacando en particular el procesamiento de señales en tiempo discreto.
- El capítulo 7 presenta la *modulación de espectro disperso*. A diferencia de las formas de modulación tradicionales que se presentan en los capítulos anteriores, el ancho de banda del canal se sacrifica a propósito en este tipo de modulación en aras de la seguridad o la protección contra *señales de interferencia*. Se explican las formas de secuencia directa y de salto de frecuencia de la modulación por espectro disperso.
- El capítulo 8 aborda las *comunicaciones por radio multiusuario*, donde una multitud de usuarios tienen acceso a un canal de radio común. Este tipo de canal de comunicación lo representan perfectamente los sistemas de comunicación por satélite e inalámbricos que se estudian aquí. Incluye una presentación del análisis del presupuesto del enlace, destacando los conceptos relacionados de antena y propagación, así como los cálculos del ruido.

- El capítulo 9 formula los *límites fundamentales en la teoría de la información*, los cuales se incorporan en los teoremas de Shannon para la compactación, la compresión y la transmisión de datos. Estos teoremas proporcionan límites superiores sobre el desempeño de las fuentes de información y los canales de comunicación. Dos conceptos fundamentales para la formulación de los teoremas son 1) la *entropía* de una fuente (cuya definición es análoga a la de la entropía en termodinámica) y 2) la capacidad del canal.
- El capítulo 10 trata la *codificación de control de errores*, la cual comprende técnicas para la codificación y la decodificación de flujos de datos digitales para su transmisión confiable por canales ruidosos. Se exponen cuatro tipos de codificación de control de errores:
  i) *Códigos de bloque lineales*: se describen completamente mediante conjuntos de palabras de código linealmente independientes, cada una de ellas consta de bits de mensaje y de bits de verificación de paridad. Estos últimos se incluyen con el fin de controlar los errores.
  ii) *Códigos cíclicos*: forman una subclase de los códigos de bloques lineales.
  iii) *Códigos convolucionales*: implican la operación en forma continua y de manera serial sobre la secuencia de mensaje.
  iv) *Códigos turbo*: proporcionan un método novedoso para la construcción de buenos códigos que se aproximan a la capacidad de Shannon del canal de una manera realizable físicamente.

  Se analizan los métodos para la generación de estos códigos, así como su decodificación.

El libro contiene también material suplementario en la forma de seis apéndices:

- En el apéndice 1 se repasa la teoría de la probabilidad.
- En el apéndice 2, acerca de la representación de señales y sistemas, se repasa la transformada de Fourier y sus propiedades, las diversas definiciones del ancho de banda, la transformada de Hilbert y los equivalentes pasobajas de las señales y sistemas de banda angosta.
- En el apéndice 3 se presenta un estudio introductorio de la función de Bessel y de su forma modificada. Las funciones de Bessel surgen en el estudio de la modulación en frecuencia, la detección no coherente de señales en ruido y la sincronización de la temporización del símbolo.
- En el apéndice 4 se introduce la función hipergeométrica confluente, cuya necesidad surge de los problemas de detección en ruido de la envolvente de señales moduladas en amplitud.
- En el apéndice 5 se ofrece una introducción a la criptografía, la cual es fundamental para las comunicaciones seguras.
- En el apéndice 6 se incluyen 12 útiles tablas de diversos tipos.

Como se mencionó antes, el objetivo primario de este libro es ofrecer un tratamiento moderno de los sistemas de comunicación adecuado para utilizarse en un curso universitario en el nivel avanzado de uno o dos semestres. La selección del material para el curso depende naturalmente de los antecedentes de los estudiantes y el interés de los maestros participantes. El material que se cubre en el libro es lo bastante extenso y profundo para satisfacer una diversidad de antecedentes e intereses, lo que permite una gran flexibilidad en la elección del contenido del curso. Como apoyo para el profesor, en el sitio del libro en la red es posible obtener un detallado manual de soluciones de todos los problemas del libro.

Agradecimientos

Deseo expresar mi profunda gratitud a los doctores Gregory J. Pottie (Universidad de California, Los Ángeles), Santosh Venkatesh (Universidad de Pennsilvania), Stephen G. Wilson (Univer-

sidad de Virginia), Gordon Stüber (Instituto de Tecnología de Georgia), Venugopal Veeraralli (Universidad de Cornell) y Granville E. Ott (Universidad de Texas en Austin) por sus revisiones críticas a una versión anterior del manuscrito y por sus numerosas sugerencias que me ayudaron a dar la forma presente al libro. El tratamiento del efecto del ruido sobre la detección de la envolvente que se presenta en el capítulo 2, se basa en las notas de un curso que puso a mi disposición el doctor Santosh Venkatesh, con quien estoy en deuda. Doy las gracias al doctor Gordon Stüber por autorizar la reproducción de la figura 6.32.

Agradezco al doctor Michael Moher (Communications Research Centre, Ottawa) la lectura de cinco capítulos de una versión anterior del manuscrito y por sus comentarios constructivos con respecto a los códigos turbo. Igualmente estoy agradecido con el doctor Brendan Frey (Universidad de Waterloo, Ontario) por su invaluable ayuda en la afinación del material sobre los códigos turbo, sus comentarios relativos a los códigos de verificación de paridad de baja densidad, el software que proporcionó para dibujar la figura 9.18 y por autorizar la reproducción de las figuras 10.27 y 10.33. Mi agradecimiento para el doctor David Conn (Universidad McMaster, Ontario) por su lectura crítica del capítulo Antecedentes y preliminares, y por sus recomendaciones acerca de cómo mejorar el material presentado ahí.

También quiero dar las gracias a los doctores Jean-Jacque Werner (Lucent Technologies, Holmdel), James Mazo (Lucent Technologies, Murray Hill), Andrew Viterbi (Qualcom, San Diego), Radford Neal (Universidad de Toronto, Ontario), Yitzhak (Irwin) Kalet (Technion, Israel), Walter Chen (Motorola), John Cioffi (Universidad de Stanford), Jon Mark (Universidad de Waterloo, Ontario) y Robert Dony (Universidad de Guelph, Ontario); les agradezco sus valiosos comentarios sobre las secciones escogidas en el libro. Aprecio también las correcciones y sugerencias para mejorar el libro del doctor Donald Wunsch II (Universidad de Missouri).

Estoy agradecido con mi estudiante de posgrado Mathini Sellathurai (Universidad McMaster) por realizar los experimentos de computadora en el libro, y con Hugh Pasika (Universidad McMaster) por muchos comentarios útiles sobre el capítulo Antecedentes y preliminares y por efectuar los cálculos relativos a algunas gráficas del libro. Aprecio sobremanera las pruebas de lectura finales que llevaron a cabo Mathini Sellathurai y Alpesh Patel.

En particular agradezco a mi editor en Wiley, Bill Zobrist, por su amplio apoyo y ayuda a lo largo de la escritura del libro. Estoy en deuda con Monique Calello, editor de producción en jefe en Wiley, por su esfuerzo incansable en la supervisión de la producción del libro en sus distintas fases. Doy las gracias a Katherine Hepburn por la publicidad y la comercialización del libro. Gracias también a Karen Tongish por su corrección cuidadosa del manuscrito, a Katrina Avery por la lectura de las pruebas finales y a Kristen Maus por elaborar el índice de la obra.

Por último, aunque no menos importante, como siempre, mi gratitud para mi coordinadora técnica, Lola Brooks, por su innagotable empeño en la mecanografía del manuscrito del libro. Igualmente, quiero expresar mi aprecio a Brigitte Maier, asistente bibliotecaria, y a Regina Bendig, especialista en información bibliográfica en la Universidad McMaster, por ayudarme en numerosas ocasiones en la búsqueda de las referencias para la bibliografía.

Simon Haykin
Ancaster, Ontario

CONTENIDO

ANTECEDENTES Y PRELIMINARES 1

1. El proceso de comunicación 1
2. Recursos de comunicación primarios 3
3. Fuentes de información 3
4. Redes de comunicación 10
5. Canales de comunicación 15
6. Proceso de modulación 19
7. Tipos de comunicación analógica y digital 21
8. Teorema de la capacidad de información de Shannon 23
9. Un problema de comunicación digital 24
10. Notas históricas 26

Notas y bibliografía 29

CAPÍTULO 1 *Procesos aleatorios* 31

1.1 Introducción 31
1.2 Definición matemática de un proceso aleatorio 32
1.3 Procesos estacionarios 33
1.4 Funciones de la media, de correlación y de covarianza 35
1.5 Procesos ergódicos 41
1.6 Transmisión de un proceso aleatorio a través de un filtro lineal invariante con el tiempo 42
1.7 Densidad espectral de potencia 44
1.8 Proceso gaussiano 54
1.9 Ruido 58
1.10 Ruido de banda angosta 64
1.11 Representación del ruido de banda angosta en términos de las componentes en fase y en cuadratura 64
1.12 Representación del ruido de banda angosta en términos de las componentes de la envolvente y de la fase 67
1.13 Onda seno más ruido de banda angosta 69
1.14 Experimentos de computadora: canal de desvanecimiento plano 71

1.15 Resumen y discusión 75
Notas y bibliografía 77
Problemas 78

CAPÍTULO 2 *Modulación de onda continua* **88**

2.1 Introducción 88
2.2 Modulación de amplitud 90
2.3 Esquemas de modulación lineal 93
2.4 Traslación de frecuencia 103
2.5 Multiplexado por división de frecuencia 105
2.6 Modulación angular 107
2.7 Modulación de frecuencia 109
2.8 Efectos no lineales en sistemas de FM 126
2.9 Receptor superheterodino 128
2.10 Ruido en sistemas de modulación de onda continua 130
2.11 Ruido en receptores lineales al utilizar detección coherente 132
2.12 Ruido en receptores de AM al utilizar detección de envolvente 135
2.13 Ruido en receptores de FM 142
2.14 Experimentos de computadora: malla de fase encadenada 157
2.15 Resumen y discusión 162
Notas y bibliografía 165
Problemas 166

CAPÍTULO 3 *Modulación por pulsos* **183**

3.1 Introducción 183
3.2 Proceso de muestreo 184
3.3 Modulación por amplitud de pulsos 188
3.4 Otras formas de modulación por pulsos 191
3.5 Compromiso ancho de banda-ruido 193
3.6 Proceso de cuantización 193
3.7 Modulación por codificación de pulsos 201
3.8 Consideraciones de ruido en sistemas PCM 209
3.9 Multiplexado por división de tiempo 211
3.10 Multiplexores digitales 214
3.11 Virtudes, limitaciones y modificaciones de la PCM 209
3.12 Modulación delta 218
3.13 Predicción lineal 223
3.14 Modulación diferencial por codificación de pulsos 227
3.15 Modulación diferencial adaptable por codificación de pulsos 229

3.16 Experimento de computadora: modulación delta adaptable 232
3.17 Estándar de codificación de audio MPEG 234
3.18 Resumen y discusión 236
Notas y bibliografía 238
Problemas 239

CAPÍTULO 4 *Transmisión de pulsos en banda base* 247

4.1 Introducción 247
4.2 Filtro de acoplamiento 248
4.3 Tasa de error debido al ruido 253
4.4 Interferencia de intersímbolo 259
4.5 Criterio de Nyquist para la transmisión binaria sin distorsión en banda base 261
4.6 Codificación de nivel correlativa 267
4.7 Transmisión PAM M-aria de banda base 275
4.8 Líneas digitales de suscriptor 277
4.9 Receptor lineal óptimo 282
4.10 Ecualización adaptable 287
4.11 Experimento de computadora: patrones del ojo 293
4.12 Resumen y discusión 296
Notas y bibliografía 297
Problemas 300

CAPÍTULO 5 *Análisis espacial de señales* 309

5.1 Introducción 309
5.2 Representación geométrica de señales 311
5.3 Conversión del canal AWGN continuo en un canal vectorial 318
5.4 Funciones de verosimilitud 322
5.5 Detección coherente de señales en ruido: decodificación de máxima verosimilitud 322
5.6 Receptor de correlación 326
5.7 Probabilidad de error 328
5.8 Resumen y discusión 337
Notas y bibliografía 337
Problemas 338

CAPÍTULO 6 *Transmisión de datos pasobanda* 344

6.1 Introducción 344
6.2 Modelo de transmisión pasobanda 348
6.3 Corrimiento de fase por llaveo coherente 349

6.4 Esquemas de modulación híbridos de amplitud/fase 368
6.5 Corrimiento de frecuencia binaria por llaveo coherente 380
6.6 Detección de señales con fase desconocida 403
6.7 Modulación ortogonal no coherente 407
6.8 Corrimiento de frecuencia binaria por llaveo no coherente 413
6.9 Corrimiento de fase por llaveo diferencial 414
6.10 Comparación de esquemas de modulación digital utilizando una sola portadora 417
6.11 Módems en la banda de voz 420
6.12 Modulación multicanal 431
6.13 Multitonos discretos 440
6.14 Sincronización 448
6.15 Experimentos de computadora: recuperación de portadora y temporización del símbolo 458
6.16 Resumen y discusión 464
Notas y bibliografía 465
Problemas 468

CAPÍTULO 7 *Modulación por espectro disperso* **479**

7.1 Introducción 479
7.2 Secuencias de pseudo ruido 480
7.3 Una noción del espectro disperso 488
7.4 Espectro disperso de secuencia directa con corrimiento de fase por llaveo binaria coherente 490
7.5 Dimensionalidad del espacio de señales y ganancia del procesamiento 493
7.6 Probabilidad de error 497
7.7 Espectro disperso de salto de frecuencia 499
7.8 Experimentos de computadora: longitud máxima y códigos de Gold 505
7.9 Resumen y discusión 508
Notas y bibliografía 509
Problemas 509

CAPÍTULO 8 *Comunicaciones por radio multiusuario* **512**

8.1 Introducción 512
8.2 Técnicas de acceso múltiple 513
8.3 Comunicaciones por satélite 514
8.4 Análisis del enlace de radio 517
8.5 Comunicaciones inalámbricas 529
8.6 Caracterización estadística de canales de multitrayectoria 535
8.7 Señalización binaria sobre un canal con desvanecimiento de Rayleigh 542
8.8 Sistemas de comunicación inalámbricos TDMA y CDMA 547
8.9 Codificación de la fuente del habla para las comunicaciones inalámbricas 550

8.10 Arreglos de antenas adaptables para las comunicaciones inalámbricas 553
8.11 Resumen y discusión 559
Notas y bibliografía 560
Problemas 562

CAPÍTULO 9 *Límites fundamentales en la teoría de la información* **567**

9.1 Introducción 567
9.2 Incertidumbre, información y entropía 568
9.3 Teorema de codificación de fuente 574
9.4 Compactación de datos 575
9.5 Canales sin memoria discretos 581
9.6 Información mutua 584
9.7 Capacidad del canal 587
9.8 Teorema de codificación del canal 589
9.9 Entropía diferencial e información mutua para conjuntos continuos 593
9.10 Teorema de la capacidad de información 597
9.11 Implicaciones del teorema de la capacidad de información 601
9.12 Capacidad de información del canal de ruido coloreado 607
9.13 Teoría de distorsión de velocidad 611
9.14 Compresión de datos 614
9.15 Resumen y discusión 616
Notas y bibliografía 617
Problemas 618

CAPÍTULO 10 *Codificación de control de errores* **626**

10.1 Introducción 626
10.2 Canales discretos sin memoria 629
10.3 Códigos de bloque lineales 632
10.4 Códigos cíclicos 641
10.5 Códigos convolucionales 654
10.6 Decodificación de máxima verosimilitud de códigos convolucionales 660
10.7 Modulación por codificación de enramado 668
10.8 Códigos turbo 674
10.9 Experimento de computadora: decodificación turbo 682
10.10 Códigos de verificación de paridad de baja densidad 683
10.11 Códigos irregulares 691
10.12 Resumen y discusión 693
Notas y bibliografía 694
Problemas 696

APÉNDICE 1 Teoría de la probabilidad 703

APÉNDICE 2 Representación de señales y sistemas 715

APÉNDICE 3 Funciones de Bessel 735

APÉNDICE 4 Funciones hipergeométricas confluentes 740

APÉNDICE 5 Criptografía 742

APÉNDICE 6 Tablas 761

GLOSARIO **771**

BIBLIOGRAFÍA **777**

ÍNDICE **792**

ANTECEDENTES Y PRELIMINARES

El material de antecedentes y preliminares que se presenta en esta sección prepara el escenario para un tratamiento estadístico de los sistemas de comunicaciones en capítulos subsecuentes. En particular describe lo siguiente:

- *El proceso de comunicación.*
- *Recursos de comunicación primarios, a saber, potencia transmitida y ancho de banda del canal.*
- *Fuentes de información.*
- *Los dos tipos primarios de conmutación: conmutación de circuitos y conmutación de paquetes.*
- *Canales de comunicación para la transportación de señales que contienen información del transmisor al receptor.*
- *El proceso de modulación, el cual es básico para los sistemas de comunicación.*
- *Sistemas de comunicación de tipos analógico y digital.*
- *Teorema de capacidad de información de Shannon.*
- *Un problema de comunicaciones digitales.*

El capítulo concluye con algunas notas históricas, como fuente de estímulo para el lector.

1. *El proceso de comunicación*[1]

En la actualidad, la *comunicación* participa en nuestra vida cotidiana de muchas diferentes maneras y por ello es fácil pasar por alto la multitud de sus facetas. Los teléfonos en nuestras manos, los radios y televisiones en nuestras habitaciones, las terminales de computadora con acceso a Internet en nuestras oficinas y hogares, así como nuestros periódicos, son capaces de proporcionar comunicaciones rápidas desde cada rincón del planeta. Las comunicaciones indican el rumbo de los barcos en altamar, los aviones en vuelo y los cohetes y satélites en el espacio. Asimismo, mantienen a un conductor en contacto con su oficina o casa a kilómetros de distancia a través del teléfono inalámbrico y se encargan de informar a un pronosticador del tiempo acerca de las condiciones que miden multitud de sensores. En realidad, la lista de aplicaciones que implican el uso de las comunicaciones de una forma u otra es casi interminable.

En sentido fundamental, la comunicación implica en forma manifiesta la transmisión de *información* de un punto a otro a través de una sucesión de procesos, como se describe aquí:

1. La generación de una *señal del mensaje*: voz, música, imagen o datos de computadora.
2. La descripción de esa señal del mensaje con cierta medida de precisión, mediante un conjunto de *símbolos*: eléctricos, auditivos o visuales.
3. La *codificación* de estos símbolos en una forma que sea adecuada para la transmisión por un medio físico de interés.
4. La *transmisión* de los símbolos codificados al destino deseado.
5. La *decodificación* y la *reproducción* de los símbolos originales.
6. La *recreación* de la señal del mensaje original, con una degradación definible en la calidad; la degradación la ocasionan las imperfecciones del sistema.

Existen, desde luego, muchas otras formas de comunicación que no implican directamente la mente humana en tiempo real. Por ejemplo, en las *comunicaciones por computadora* que comprenden dos o más computadoras que se comunican entre ellas, las decisiones humanas sólo participan en la formulación de los programas o comandos de la máquina o en la supervisión de los resultados.

Independientemente de la forma del proceso de comunicación que se esté considerando, existen tres elementos básicos para todo sistema de comunicación: *transmisor, canal* y *receptor,* como se indica en la figura 1. El transmisor se localiza en un punto en el espacio, el receptor se ubica en algún otro punto separado del transmisor, y el canal es el medio físico que los conecta. El propósito del transmisor es convertir la *señal del mensaje* producida por la *fuente de información* en una forma adecuada para la transmisión por el canal. Sin embargo, como la señal transmitida se propaga a lo largo del canal, se distorsiona debido a imperfecciones de este último. Aún más, el ruido y las señales de interferencia (originadas de otras fuentes) se suman a la salida del canal, con el resultado de que la *señal recibida* es una versión corrompida de la *señal transmitida*. El receptor tiene la tarea de actuar sobre la señal recibida de manera que reconstruya para el usuario una forma reconocible de la señal del mensaje original.

Existen dos modos básicos de comunicación:

1. *Transmisión*, que implica el uso de un transmisor poderoso y de numerosos receptores cuya construcción es relativamente económica. Aquí las señales que contienen la información sólo fluyen en una dirección.
2. *Comunicación punto a punto*, en la cual los procesos de comunicación se llevan a cabo por un enlace entre un solo transmisor y un receptor. En este caso, casi siempre existe un flujo bidireccional de las señales que llevan la información que requiere el uso de un transmisor y un receptor en cada extremo del enlace.

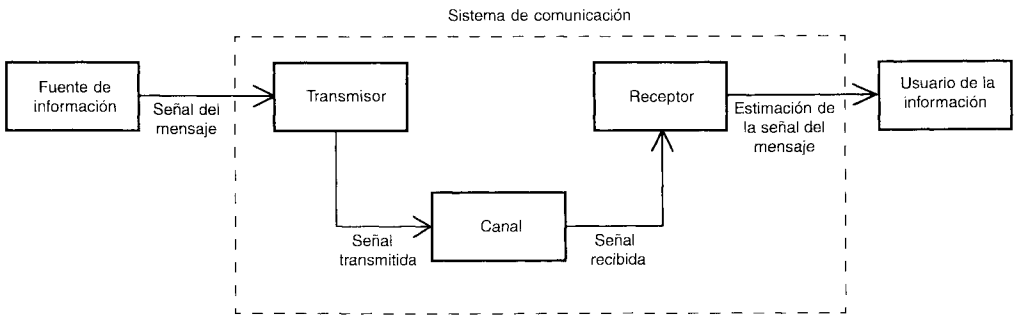


FIGURA 1 Elementos de un sistema de comunicación.

El modo de comunicación por transmisión se ejemplifica a través de la radio y la televisión; el ubicuo teléfono constituye el medio para una forma de comunicación punto a punto. Otro ejemplo de la comunicación punto a punto es el enlace entre una estación en Tierra y un robot que navega en la superficie de un planeta distante.

Todos estos diferentes sistemas de comunicación, así como otros que no se mencionan aquí, comparten un rasgo común: el proceso de comunicación fundamental en todos y cada uno de ellos es de naturaleza *estadística*. En realidad, a esta importante razón se debe que gran parte de este libro se dedique a los fundamentos estadísticos de los sistemas de comunicación. De ese modo, desarrollamos una exposición de los aspectos más importantes involucrados en el estudio de las diferentes metodologías de comunicación y proporcionamos un foro natural para sus evaluaciones comparativas.

2. *Recursos de comunicación primarios*

En un sistema de comunicación se emplean dos recursos primarios: la *potencia transmitida* y el *ancho de banda del canal*. La primera corresponde a la potencia promedio de la señal transmitida. El ancho de banda del canal se define como la banda de frecuencias destinadas a la transmisión de la señal del mensaje. Un objetivo de diseño general del sistema es utilizar estos dos recursos en la forma más eficiente posible. En la mayor parte de los canales de comunicación, uno de los recursos quizá se considere más importante que el otro. En consecuencia, es posible clasificar los canales de comunicación como *limitados en potencia* o *limitados en banda*. Por ejemplo, el circuito telefónico es un típico canal limitado en banda, en tanto que los enlaces de comunicación espaciales o el canal de satélite son comúnmente limitados en potencia.

Cuando el espectro de una señal del mensaje se extiende hasta frecuencias de cero o bajas, definimos el ancho de banda de la señal como aquella frecuencia superior por encima de la cual es despreciable el contenido espectral de la señal y, por tanto, innecesario para transmitir información. Por ejemplo, el espectro de voz promedio se extiende más allá de 10 kHz, aunque la mayor parte de la potencia promedio se concentra en el intervalo de 100 a 600 Hz, y una banda de 300 a 3100 Hz proporciona buena articulación. De acuerdo con lo anterior, encontramos que los circuitos telefónicos que responden bien a este último intervalo de frecuencias proporcionan un servicio telefónico comercial satisfactorio.

Otro importante punto que debemos tener presente es la inevitable presencia de ruido en un sistema de comunicación. El *ruido* se refiere a ondas indeseables que tienden a perturbar la transmisión y el procesamiento de las señales del mensaje en un sistema de comunicación. La fuente de ruido puede ser interna o externa al sistema.

Una forma cuantitativa de tener en cuenta el efecto del ruido consiste en introducir la *relación señal a ruido* (SNR, por sus siglas en inglés) como un parámetro del sistema. Por ejemplo, es posible definir dicha relación en la entrada del receptor como *la proporción entre la potencia promedio de la señal y la potencia promedio del ruido*, midiéndose ambas en el mismo punto. La práctica común es expresar la SNR en *decibeles* (dB), definida como 10 veces el logaritmo decimal (base 10) de la proporción de potencia. Por ejemplo, las relaciones de señal a ruido de 10, 100 y 1000 corresponden, respectivamente, a 10, 20 y 30 dB.

3. *Fuentes de información*

El ambiente de las telecomunicaciones está dominado por cuatro importantes fuentes de información: *habla*, *música*, *imágenes* y *datos de computadora*. Es posible caracterizar una fuente de información en términos de la señal que porta la información. Una *señal* se define como una función de un solo valor de *tiempo* que desempeña el papel de la variable independiente; en cualquier instante, la función tiene un valor único. La señal puede ser *unidimensional*, como en el caso del habla, la música o los datos de computadora; *bidimensional*, como en el caso de las imágenes; *tridimensional*,

como en el caso de los datos de video; y *cuadrimensional*, como en el caso de datos de volumen a lo largo del tiempo. En lo que sigue explicaremos las diferentes fuentes de información.

i) El *habla* es el método fundamental de la comunicación humana. Específicamente, el proceso de comunicación del habla implica la transferencia de información de un hablante a un escucha, lo cual ocurre en tres etapas sucesivas:

- *Producción*. Un mensaje en la mente del hablante que quiere expresarse se representa mediante una *señal del habla* que está compuesta por sonidos (es decir, ondas de presión) generados dentro del tracto vocal y cuyo arreglo está gobernado por las reglas del lenguaje.
- *Propagación*. Las ondas sonoras se propagan por el aire a una velocidad de 300 m/s, llegando a los oídos del escucha.
- *Percepción*. El oyente descifra los sonidos que llegan en un mensaje recibido, completando de ese modo la cadena de acontecimientos que culminan en la transferencia de información del hablante al escucha.

El proceso de producción del habla puede considerarse como una forma de filtrado, en el cual una *fuente sonora excita a un filtro del tracto vocal*. El tracto vocal se compone de un tubo de área de sección transversal no uniforme, que se inicia en la *glotis* (es decir, la abertura entre las cuerdas vocales) y que termina en el *labio*. Cuando el sonido se propaga a lo largo del tracto vocal, el espectro (es decir, el contenido de frecuencia) es formado por la selectividad de frecuencia del tracto vocal; este efecto es un poco similar al fenómeno de resonancia que se observa en los tubos de órgano. El punto importante a observar aquí es que el espectro de potencia (es decir, la distribución de la potencia promedio de largo plazo en función de la frecuencia) del habla tiende a cero para frecuencia cero y alcanza un máximo en la vecindad de unos cuantos cientos de hertz. Sin embargo, para ubicar este asunto en la perspectiva adecuada, debemos recordar que el mecanismo de audición resulta muy sensible a la frecuencia. Asimismo, el tipo de sistema de comunicación que se está considerando tiene una presencia importante en la banda de frecuencias que se considera "esencial" para el proceso de comunicación. Por ejemplo (como se mencionó antes), un ancho de banda de 300 a 3100 Hz se considera adecuado para la comunicación telefónica comercial.

ii) La segunda fuente de información, la *música*, se origina de instrumentos como el piano, el violín y la flauta. Las notas que produce un instrumento musical quizá duren un breve intervalo de tiempo, como cuando se presiona una tecla en un piano, o es posible sostenerlas durante un intervalo de tiempo largo, como en el caso de un flautista que sostiene una nota prolongada. Por lo común, la música tiene dos estructuras: una *melódica*, compuesta por una secuencia de tiempo de sonidos, y una *armónica*, que se compone de un conjunto de sonidos simultáneos. Al igual que la señal del habla, una señal musical es bipolar. Sin embargo, difiere de la del habla en que su espectro ocupa una banda mucho más ancha de frecuencias que se puede extender hasta cerca de 15 kHz. De acuerdo con lo anterior, las señales musicales demandan un ancho de banda de canal mucho más amplio que las señales del habla para su transmisión.

iii) La tercera fuente de información, las *imágenes*, se perciben por medio del sistema visual humano. La imagen puede ser *dinámica*, como en la televisión, o *estática*, como en el facsímil. Se considera primero el caso de la televisión, en el cual las imágenes en movimiento se convierten en señales eléctricas para facilitar su transporte desde el transmisor hasta el receptor. Para hacerlo de ese modo, cada imagen completa se *escanea (explora ópticamente) en forma secuencial*. El proceso de escaneo se realiza en una *cámara de TV*. En la televisión en blanco y negro, la cámara contiene una óptica diseñada para enfocar una imagen sobre un *fotocátodo* compuesto por un gran número de elementos fotosensibles. El patrón de carga generado de esa manera sobre la superficie fotosensible se escanea mediante un *haz electrónico*, produciéndose en esa forma una corriente de salida que varía *temporalmente* con la manera en la cual la brillantez de la imagen original varía *espacialmente* de un punto a otro. La corriente de salida

resultante se denomina *señal de video*. El tipo de escaneo que se utiliza en la televisión es una forma de muestreo espacial que se denomina *escaneo de trama*, el cual convierte la intensidad de una imagen bidimensional en una forma de onda unidimensional; esto se asemeja un poco a la manera en la cual leemos un papel impreso, ya que el escaneo se realiza de izquierda a derecha en un patrón de línea por línea. En la televisión analógica de Estados Unidos, una imagen se divide en 525 líneas, lo cual constituye una *trama*. Cada trama se divide en dos *campos entrelazados* y cada uno de ellos consta de 262.5 líneas. Para facilitar la exposición, nos referiremos a los dos campos como I y II. El procedimiento de escaneo se ilustra en la figura 2. Las líneas del campo I se describen como líneas continuas y las relativas al campo II se presentan como líneas discontinuas. El *inicio* y *fin* de cada campo se incluye también en la figura. El campo I se escanea primero. El punto de escaneo de la cámara de TV se mueve con velocidad constante a lo largo de cada línea del campo de izquierda a derecha y se mide la intensidad de la imagen en el centro del punto; el punto de escaneo por sí mismo es parcialmente responsable del promedio espacial local de la imagen. Cuando se alcanza el extremo de una línea particular, el punto de escaneo regresa rápidamente (en una dirección horizontal) al principio de la siguiente línea de abajo en el campo. Este retroceso se denomina *retrazado horizontal*. El proceso de exploración que acaba de describirse continúa hasta cubrir el campo completo. Al llegar a esta condición, el punto de escaneo se mueve rápidamente (en dirección vertical) desde el final del campo I hasta el inicio del campo II. Este segundo retroceso se denomina *retrazado vertical*. El campo II se trata de la misma manera que el campo I. El tiempo que toma la exploración de cada campo es igual a 1/60 s. En forma correspondiente, el tiempo que se necesita para explorar una trama o una imagen completa equivale a 1/30 s. Con 525 líneas en una trama, la *frecuencia de exploración de líneas* es igual a 15.75 kHz. De esta manera, al mostrar 30 imágenes fijas por segundo en el tubo de imagen de un receptor de TV, el ojo humano las percibe como imágenes en movimiento. Este efecto se debe a un fenómeno conocido como la *persistencia de la visión*. Durante los intervalos de retrazado horizontal y vertical, el tubo de imagen se hace inoperativo por medio de *pulsos de anulación* o *blanqueo* que se generan en el transmisor. Además, la sincronización entre las diversas operaciones de exploración tanto en el transmisor como en el receptor se consiguen mediante pulsos especiales que se transmiten durante los periodos de anulación; de esta manera, los pulsos de sincronización no se muestran sobre la imagen reproducida. La calidad de reproducción de la imagen de TV está limitada por dos factores básicos:

1. El número de líneas disponibles en una exploración (escaneo) de trama, que limita la resolución de la imagen en la dirección vertical.
2. El ancho de banda del canal disponible para transmitir la señal de video, que limita la resolución de la imagen en la dirección horizontal.

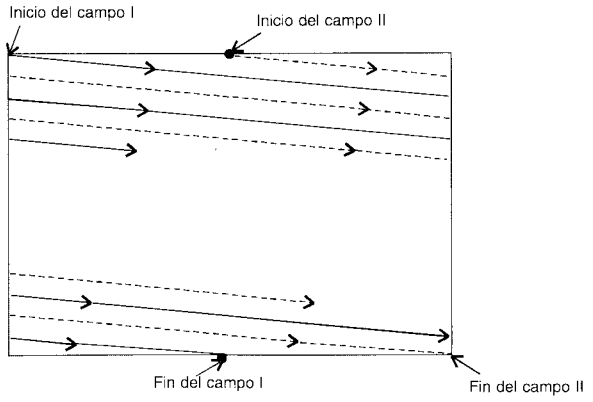


FIGURA 2 Exploración óptica (escaneo) de trama entrelazada.

En cada dirección, la *resolución* se expresa en términos del número máximo de líneas que se alternan entre el negro y el blanco y que pueden determinarse en la imagen de TV junto con la dirección pertinente de un observador humano. En el sistema NTSC (siglas del National Television System Committee de Estados Unidos), que constituye el estándar en ese país, los valores de los parámetros que se utilizan producen un *ancho de banda de video* de 4.2 MHz, el cual se extiende hasta la frecuencia cero. Este ancho de banda es varias órdenes de magnitud más grande que el de la señal del habla. Adviértase, asimismo, que en tanto que una señal de voz es bipolar, una de video (televisión) es inherentemente positiva (es decir, unipolar).

En la TV de color, la percepción del color se basa en tres tipos de receptores de color (conos) en el ojo humano: rojo, verde y azul, cuyas longitudes de onda son 570 nm, 535 nm y 445 nm, respectivamente. Estos tres colores se conocen como los *colores primarios*, ya que cualquier otro color que se encuentra en la naturaleza se aproxima por medio de una mezcla aditiva de ellos. Esta realidad física constituye la base para la transmisión del color en las emisiones de la televisión comercial. Los tres colores primarios se representan mediante las señales de video $m_R(t)$, $m_G(t)$ y $m_B(t)$, respectivamente. Para conservar el ancho de banda y producir una imagen que es posible ver sobre un receptor de televisión convencional en blanco y negro (monocromático), la transmisión de estos tres colores primarios se consigue observando que éstos pueden representarse en forma única por medio de tres señales que son combinaciones lineales independientes de $m_R(t)$, $m_G(t)$ y $m_B(t)$. Las tres señales son como sigue:

- Una *señal de luminancia*, $m_L(t)$, que produce una versión en blanco y negro de la imagen en color cuando ésta se recibe en un receptor de televisión convencional monocromático.
- Un par de señales $m_I(t)$ y $m_Q(t)$, denominadas las *señales de crominancia*, indican la forma en que el color de la imagen se desvía de las sombras de grises.

A la señal de luminancia $m_L(t)$ se le asigna el ancho de banda completo de 4.2 MHz. Debido a ciertas propiedades de la visión humana, las pruebas demuestran que si los anchos de banda nominales de las señales de crominancia $m_I(t)$ y $m_Q(t)$ son 1.6 MHz y 0.6 MHz, respectivamente, es posible la reproducción satisfactoria del color.

Si se considera a continuación una *máquina de facsímil* (fax), se sabe que el propósito de ésta es transmitir imágenes fijas por un canal de comunicación (principalmente telefónico). Una máquina de este tipo constituye una instalación muy popular para la transmisión de texto escrito a mano o impreso de un punto a otro; la transmisión de texto mediante facsímil se considera simplemente como la transmisión de una imagen. El principio básico que se emplea para la generación de la señal en una máquina de facsímil consiste en explorar ópticamente un documento original (imagen) y utilizar un sensor de imágenes para convertir la luz en una señal eléctrica.

iv) Por último, las *computadoras personales* (PC) se han vuelto parte integral de nuestra vida diaria. Las utilizamos para correo electrónico, intercambio de software y para compartir recursos. El texto que se transmite por medio de una PC suele codificarse utilizando el *código estándar estadounidense para el intercambio de información* (ASCII, por sus siglas en inglés), que es el primer código que se desarrolló específicamente para las comunicaciones por computadora. Cada carácter en ASCII se representa por medio de siete *bits de datos* que constituyen un patrón binario único conformado por ceros y unos; bit es el acrónimo en inglés para dígito binario. De esta manera, un total de $2^7 = 128$ caracteres diferentes pueden representarse en ASCII. Los caracteres corresponden a diversas letras minúsculas y mayúsculas, números, símbolos de control especiales y de puntuación utilizados comúnmente, como @, $ y %. Algunos de los símbolos de "control" especiales, como BS (*backspace* o retroceso) y CR (*carriage return* o retorno de carro), se utilizan para controlar la impresión de caracteres sobre una página. Otros símbolos, como ENQ (*enquiry* o pregunta) y ETB (*end of transmission block* o fin del bloque de transmisión), se utilizan con fines de comunicación. (En la tabla A6.1 se proporciona una lista completa de caracteres ASCII.) Los siete bits de datos se

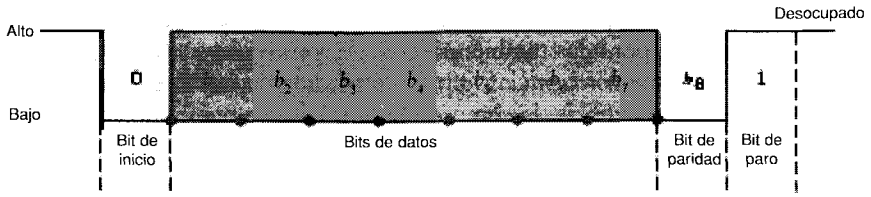


FIGURA 3 El formato de bit para enviar datos en serie asíncronos que se utiliza en el estándar RS-232.

ordenan empezando con el bit más significativo b_7 hasta el menos significativo b_1, como se ilustra en la figura 3. Al final de los bits de datos, se agrega un bit adicional b_8 con el fin de *detectar errores*. Este bit se denomina *bit de paridad*. Una secuencia de ocho bits se conoce como un *byte* o un *octeto*. El bit de paridad se establece de tal manera que el número total de unos en cada byte sea impar para paridad impar y par para paridad par. Suponga, por ejemplo, que los comunicadores concuerdan en utilizar paridad par; en ese caso el bit de paridad será un 0 cuando el número de unos en el bit de datos es par y un 1 cuando es impar. Por consiguiente, si un solo bit en un byte se recibe con error y por eso viola la regla de paridad par, es posible detectarlo y corregirlo después por medio de la retransmisión. Las computadoras personales se conectan a menudo por medio de sus puertos RS (estándar recomendado)-232. Cuando los datos ASCII (en realidad todos los datos de caracteres) se transmiten a través de estos puertos, un bit de inicio, fijado como 0, y uno o más bits de paro, fijados como 1, como se muestra en la figura 3, se agregan para proporcionar un marco para los caracteres. Cuando la transmisión está inactiva, se envía una larga serie de unos de manera que se mantenga viva la conexión del circuito. En la figura 3, los símbolos 0 y 1 se designan como "bajo" y "alto", respectivamente. Algunas veces se les llama también "espacio" y "marca"; la última terminología proviene de la época de la telegrafía. El texto que se elabora en una computadora personal suele almacenarse y luego transmitirse por un canal de comunicación (por ejemplo, un canal telefónico) con un solo carácter enviado a la vez. Esta forma de transmisión de datos recibe el nombre de *transmisión asíncrona*, en oposición a *transmisión síncrona*, en la cual una secuencia completa de caracteres codificados se envía por el canal en una transmisión larga. Los caracteres codificados producidos por una mezcla de terminales asíncronas y síncronas se combinan mediante *multiplexores de datos*. La corriente multiplexada de datos que se forma de este modo se aplica después en un dispositivo denominado *módem* (modulador-demodulador) con el fin de realizar la transmisión por el canal.

En resumen, los datos generados por computadora y las señales de televisión constituyen en ambos casos *señales de banda ancha*, porque su contenido de potencia ocupa un intervalo amplio de frecuencias. Otra característica importante de la comunicación de datos entre computadoras personales es la de *ráfaga*, lo cual significa que la información suele transmitirse de una terminal a otra en ráfagas con periodos de silencio entre ellas. En realidad, el tráfico de datos que involucra a las computadoras en una forma u otra tiende a ser de naturaleza de ráfaga. Esto contrasta con el tráfico en una red de transmisión digital en la cual la voz o el video interactivo son, en términos relativos, *continuos*.

Otra forma en la cual usamos la computadora es para *descargar* formas comprimidas de datos de texto, audio y video de un proveedor de servicios en una localidad remota. La *compresión de datos* es un medio práctico para el almacenamiento y la transmisión eficiente de estos tipos de datos. Un sistema de compresión de datos está compuesto por un *codificador* y un *decodificador*, donde se realizan, respectivamente, la compresión y reconstrucción de una corriente de datos entrantes. Básicamente, existen dos formas de compresión de datos:

1. La *compresión sin pérdida* elimina la información *redundante* contenida en los datos de interés. Se dice que la compresión es sin pérdidas debido a que es completamente

reversible, porque es posible reconstruir en forma exacta los datos originales. La compresión sin pérdidas también se conoce como *compactación de datos*.

2. La *compresión con pérdidas* implica la pérdida de información de una manera controlada; en consecuencia, la compresión quizá no sea por completo reversible. Sin embargo, este tipo de compresión es capaz de alcanzar una razón de compresión superior a la que se obtiene con los métodos sin pérdidas.

En el texto digital se requiere la compresión sin pérdidas. En este contexto mencionamos el *algoritmo de Lempel-Ziv*, que es intrínsecamente adaptable y capaz de codificar grupos de símbolos fuente que ocurren con frecuencia. Logra una compresión de aproximadamente 55% en el texto en inglés ordinario, lo que, en términos generales, corresponde a la compresión que se lograría al codificar pares de letras. El algoritmo de Lempel-Ziv es una forma de *codificación entrópica* o *fuente*, la cual se explica en el capítulo 9.

En muchas otras aplicaciones, la compresión con pérdidas suele ser el método preferido y es posible que su empleo reduzca de manera sustancial el tamaño de los datos sin alterar en forma significativa la calidad de percepción de una señal de imagen o de audio. En tal tipo de aplicaciones, resulta aceptable esta forma de compresión de datos, y constituye una necesidad en las aplicaciones de transmisión de datos de alto rendimiento como lo es Internet. Sin embargo, en algunas otras aplicaciones, como una imagen clínica, la calidad de la imagen médica (por ejemplo, radiografías de rayos X digitales) no debe degradarse en la reconstrucción.

Para que las aplicaciones de audio y video digitales que implican el almacenamiento o transmisión resulten viables en el mercado actual, necesitamos algoritmos de compresión *estándar* que permitan la interoperabilidad de equipos producidos por fabricantes diferentes. En este contexto mencionamos tres de los más prominentes algoritmos de compresión estándar que satisfacen las diferentes necesidades:

- El *estándar de codificación de imágenes JPEG*[2] está diseñado para comprimir imágenes a todo color o de escala de grises de escenas naturales del mundo real explotando las limitaciones conocidas del sistema visual humano; JPEG corresponde a las siglas en inglés de Joint Photographic Expert Group (Grupo Unido de Expertos en Fotografía). En la entrada del codificador, los *elementos de la imagen* o *pixeles* se agrupan en bloques de 8×8, los cuales se aplican a un tipo de transformada de Fourier conocida como la *transformada coseno discreta* (DCT, por sus siglas en inglés).[3] La DCT descompone cada bloque de pixeles en un conjunto de 64 coeficientes que satisfacen con mucha precisión dos objetivos relacionados:
  1. Los coeficientes deben estar lo *menos correlacionados* posible.
  2. La energía de la señal de entrada debe empaquetarse en el *menor número de coeficientes posible*.

  La siguiente operación en el codificador es la *cuantización*, donde se *redondea* cada uno de los 64 coeficientes de la DCT. En el estándar JPEG, la cuantización se lleva a cabo junto con una *tabla de cuantización* proporcionada por el usuario como una entrada para el codificador. Cada elemento de la tabla es un entero de 1 a 255 que especifica el tamaño del escalón de los coeficientes de la DCT, lo que, a su vez, permite la representación de cada coeficiente de la DCT cuantizado por medio de una palabra de código de 8 bits. Básicamente, el propósito de la cuantización es descartar información que no es discernible perceptiblemente. La cuantización constituye un *mapeo de muchos a uno* y, por tanto, es la principal fuente de pérdida en el codificador. La operación final en el codificador es la *codificación de Huffman*, una forma de codificación entrópica (fuente) que se explica también en el capítulo 9 y que consigue compresión de datos adicional sin pérdidas codificando los coeficientes de la DCT cuantizados de acuerdo con sus características estadísticas. En el decodificador, la reconstrucción de datos se lleva a cabo a través de una secuencia de operaciones que son las inversas a las del codificador,

a saber: decodificación de Huffman, cuantización inversa de acuerdo con la tabla de cuantización y, por último, la DCT inversa.

- El *estándar de codificación de video MPEG-1*[4] se diseñó fundamentalmente para comprimir señales de video a 30 tramas por segundo (cps) en una corriente de bits que corre a razón de 1.5 megabits por segundo (Mb/s); MPEG corresponde a las siglas en inglés de Grupo de Expertos de Fotografía en Movimiento. El estándar de codificación de video MPEG-1 consigue esta meta de diseño explotando cuatro formas básicas de redundancia presente inherentemente en los datos de video:
  1. Redundancia de intertrama (temporal).
  2. Redundancia interpixel dentro de una trama.
  3. Redundancia psicovisual.
  4. Redundancia de codificación entrópica.

  La explotación de la redundancia de intertrama es lo que distingue al MPEG-1 del JPEG. En principio, las tramas vecinas en secuencias de video ordinarias están altamente *correlacionadas*. El significado de esta alta correlación es que, en un sentido promedio, una señal de video no cambia rápidamente de una trama a la siguiente y, en consecuencia, la diferencia entre tramas adyacentes tiene una varianza (es decir, potencia promedio) que es mucho más pequeña que la varianza de la propia señal de video. De acuerdo con ello, la redundancia de intertrama puede reducirse en forma considerable para producir una señal de video comprimida más eficientemente. Dicha reducción se consigue a través del uso del *pronóstico* para estimar cada trama a partir de sus vecinas; el error de pronóstico que resulta se transmite para la estimación y compensación del movimiento. El pronóstico es no lineal en virtud de la naturaleza del problema. Como en el caso del estándar JPEG, la redundancia de interpixeles se reduce mediante el uso combinado de la DCT, la cuantización y la codificación entrópica sin pérdidas. El resultado final es que el video de movimiento completo se convierte en una corriente de 1.5 Mb/s de datos de computadora que es posible almacenar en discos compactos o integrar con textos y gráficas. Lo que resulta más importante es que el video de movimiento completo y el audio asociado pueden entregarse a partir de redes de computadora y de telecomunicaciones existentes, lo cual posibilita satisfacer la necesidad de *video sobre demanda* en Internet.
- El *estándar de codificación de audio MPEG-1*[5] se basa en la *codificación de percepción*, que es un proceso de preservación de forma de onda; es decir, la forma de onda de amplitud-tiempo de la señal de audio decodificada se aproxima mucho a la de la señal de audio original. En términos generales, el proceso de codificación comprende cuatro operaciones distintas:
  1. *Mapeo de tiempo-frecuencia*; por este medio la señal de audio se descompone en subbandas múltiples.
  2. *Modelación psicoacústica*, que opera en forma simultánea sobre la señal de audio de entrada para calcular ciertos umbrales utilizando reglas conocidas del comportamiento psicoacústico del sistema auditivo humano.
  3. *Cuantización y codificación*, que, junto con el modelo psicoacústico, trabaja en la salida del mapeador de tiempo-frecuencia a fin de mantener el ruido que resulta del proceso de cuantización en un nivel inaudible.
  4. *Empaque de tramas*, que se utiliza para formatear las muestras de audio cuantizadas en una corriente de bits decodificable.

  El modelo psicoacústico se conforma a partir de un fenómeno de percepción conocido como *enmascaramiento auditivo*. Específicamente, el oído humano *no* percibe el ruido de cuantización en una banda de frecuencia determinada si la potencia del ruido promedio se encuentra por debajo del *umbral de enmascaramiento* (es decir, el umbral de apenas la distorsión perceptible). Para una banda de frecuencia de interés determinada, el

umbral de enmascaramiento varía con la frecuencia a lo largo de esa banda. El *umbral de enmascaramiento mínimo* es el que se emplea en el modelo psicoacústico en una base de banda por banda. Por ejemplo, el resultado final de utilizar el estándar MPEG-1 en los dos canales de audio de un programa estereofónico es que cada señal de audio digitalizada, que se obtiene a razón de 768 kilobits por segundo (kb/s), se comprime a una velocidad tan baja como 16 kb/s. (La velocidad de los datos de entrada de 768 kb/s corresponde a una velocidad de muestreo de 48 kHz, siendo cada muestra representada por una palabra de código de 16 bits.) De este modo, el estándar de codificación de audio MPEG-1 es adecuado para el almacenamiento de señales de audio en medios económicos o su transmisión sobre canales con ancho de banda limitado, mientras mantiene al mismo tiempo la calidad de percepción.

4. *Redes de comunicación*[6]

Una *red de comunicación* (o simplemente *red*), ilustrada en la figura 4, está compuesta por una interconexión de varios *ruteadores* integrados por procesadores inteligentes (por ejemplo, microprocesadores). El propósito fundamental de estos procesadores es dirigir los datos a través de la red, de ahí su nombre. Cada ruteador tiene una o más *hosts* unidas a él. Las hosts son dispositivos que se comunican entre sí. La red se diseña para servir como recursos compartidos para datos móviles intercambiados entre hosts de una manera eficiente y para proporcionar un marco de referencia que soporte nuevas aplicaciones y servicios.

La red telefónica es un ejemplo de una red de comunicación en la que se emplea la *comunicación de circuitos* para ofrecer una trayectoria de comunicaciones dedicada, *o circuito*, entre dos hosts. El circuito consta de una secuencia conectada de enlaces de una fuente al destino. Por ejemplo, los enlaces pueden estar compuestos por ranuras de tiempo para las cuales está disponible un canal común al que tengan acceso una multitud de usuarios. El circuito, una vez establecido, se mantiene ininterrumpido durante la transmisión. La conmutación del circuito suele controlarse por medio de un mecanismo de control jerárquico centralizado con conocimiento de la organización de la red. Para establecer una conexión de conmutación de circuito, se dimensiona una trayectoria disponible a través de la red y se destina luego al uso exclusivo de las dos hosts que desean comunicarse. En particular, una señal de solicitud de llamada debe propagarse todo el tiempo hacia el destino y reconocerse antes de que sea posible el inicio de la transmisión. En esas condiciones, la red es efectivamente transparente para los usuarios. Esto significa que durante el tiempo de conexión, el ancho de banda y los recursos asignados al circuito son en esencia "poseídos" por las dos hosts hasta que el circuito se desconecta. De ese modo, el circuito representa un uso eficiente de recursos únicamente en el grado en el que el ancho de banda asignado se use con propiedad. A pesar de que la red

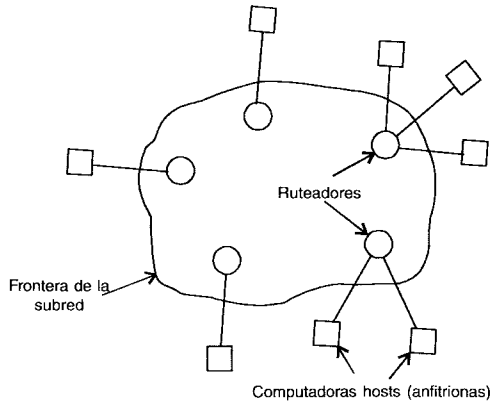


FIGURA 4 Red de comunicación.

telefónica se utiliza para transmitir datos, la voz constituye el grueso del tráfico de la red. En realidad la conmutación del circuito es bastante apropiada para la transmisión de señales de voz, ya que la voz da origen a un tráfico de flujo y las conversaciones de voz tienden a ser de larga duración (alrededor de 2 minutos en promedio), comparadas con el tiempo que se necesita para inicializar el circuito (alrededor de 0.1 a 0.5 segundos). Además, en la mayor parte de las conversaciones existe un flujo de información para un porcentaje relativamente grande del tiempo de conexión, el cual logra que la conmutación de circuitos sea la más adecuada para las conversaciones.

En la conmutación de circuitos, un enlace de comunicaciones se comparte entre las diferentes sesiones que utilizan ese enlace en un esquema de asignación *fija*. En cambio, en la *conmutación de paquete*, la compartición se realiza con base en la *demanda*, de modo que ésta tiene una ventaja sobre la conmutación de circuito en el sentido de que cuando un enlace tiene tráfico que enviar, éste quizá se utilice de manera más completa.

El principio de red de la conmutación de paquete es "almacenar y adelantar". Específicamente, en una *red de conmutación de paquetes*, cualquier mensaje mayor que un tamaño definido se subdivide antes de la transmisión en segmentos que no excedan dicho tamaño. Los segmentos se conocen comúnmente como *paquetes*. El mensaje original se reensambla en el destino paquete por paquete. Es posible considerar a la red como una bolsa distribuida de *recursos de red* (es decir, ancho de banda del canal, búfers y procesadores de conmutación) cuya capacidad se *comparte dinámicamente* mediante una comunidad de hosts en competencia que desean comunicarse. En contraste, en una red de conmutación de circuitos, los recursos se dedican a un par de hosts durante el periodo completo que se mantienen en sesión. De acuerdo con esto, la conmutación de paquetes es mucho más adecuada en un ambiente de comunicación de computadora en el que se intercambian ocasionalmente ráfagas de datos entre las hosts. Sin embargo, el uso de la conmutación de paquetes requiere poner en práctica un cuidadoso *control* en las demandas de los usuarios; de otra manera, es factible que se abuse seriamente de la red.

El diseño de una *red de datos* (es decir, una red en la cual las hosts estén integradas en su totalidad por computadoras y terminales) quizá proceda de una manera ordenada al considerar la red en términos de una *arquitectura en capas*, considerada como una jerarquía de capas anidadas. El término *capa* se refiere a un proceso o dispositivo dentro de un sistema de cómputo, diseñado para realizar una función específica. Naturalmente, los diseñadores de una capa estarán íntimamente familiarizados con sus detalles internos y operación. Sin embargo, a nivel del sistema, un usuario observa la capa meramente como una "caja negra" que se describe en términos de entradas, salidas y la relación funcional entre salidas y entradas. En una arquitectura de capas, cada una de ellas considera a la siguiente inferior como una o más cajas negras con cierta especificación funcional dada que utilizará la capa superior particular. Por consiguiente, el altamente complejo problema de comunicación en las redes de datos se resuelve como un conjunto manejable de funciones de intercierre perfectamente definido. Es esta línea de razonamiento la que ha conducido a la creación del *modelo de referencia de interconexión de sistemas abiertos* (OSI)[7] por parte de un subcomité de la International Organization for Standardization. El término *abierto* se refiere a la capacidad de cualesquiera dos sistemas que conformen el modelo de referencia y a sus estándares asociados para interconectarse.

En el modelo de referencia OSI, las funciones de comunicaciones y de conexión relativa se organizan como una serie de capas o *niveles*, con *interfaces* bien definidas y con cada capa construida sobre su predecesora. En particular, cada capa efectúa un subconjunto relacionado de funciones primitivas y se apoya en la siguiente capa inferior para realizar funciones primitivas adicionales. Además, cada una ofrece ciertos servicios a la siguiente capa superior y aísla a esta última de los detalles de implementación de esos servicios. Entre cada par de capas, existe una *interfaz*, la cual define los servicios que ofrece la capa inferior a la superior.

El modelo OSI está compuesto de siete capas, según se ilustra en la figura 5; esta figura contiene también una descripción de las funciones de las capas individuales del modelo. La capa k en el sistema A, por ejemplo, se comunica con la capa k en algún otro sistema B en

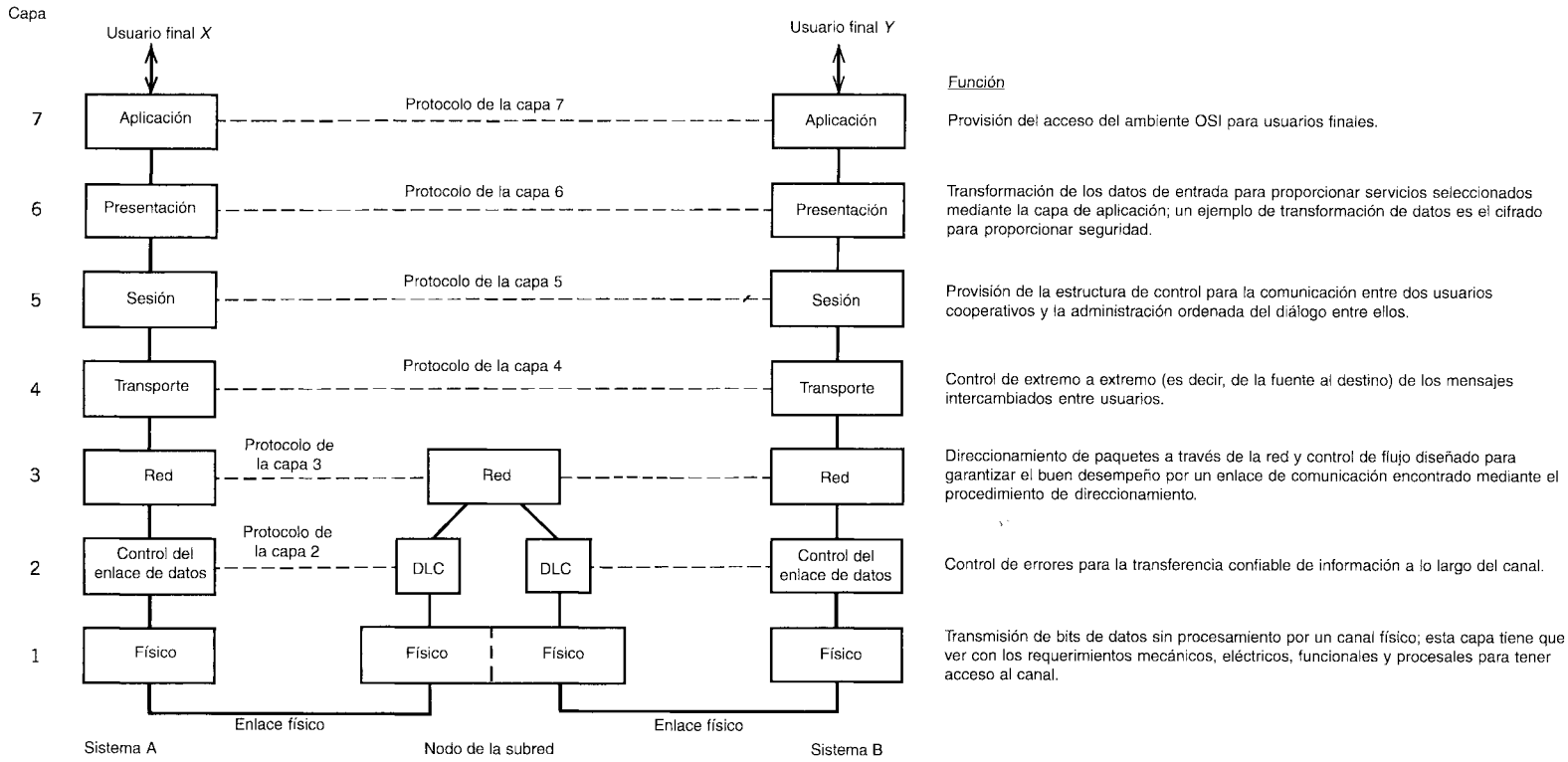


FIGURA 5 Modelo OSI; el acrónimo DLC a la mitad de la figura significa *control del enlace de datos*.

concordancia con un conjunto de reglas y convenciones, que constituyen colectivamente el *protocolo k* de la capa, donde $k = 1, 2, ..., 7$. (El término *protocolo* se ha tomado del uso común y describe al comportamiento social convencional entre seres humanos.) Las entidades que comprenden a las capas correspondientes en diferentes sistemas se denominan *procesos de iguales*. En otras palabras, la comunicación se consigue al contar con procesos de iguales en dos sistemas diferentes que se comunican por medio de un protocolo, siendo este mismo definido por medio de un conjunto de reglas o procedimientos. La comunicación física entre procesos de iguales sólo existe en la capa 1. Por otro lado, las capas de la 2 a la 7 se encuentran en *comunicación virtual* con sus iguales distantes. Sin embargo, cada una de estas seis capas tiene la posibilidad de intercambiar datos e información de control con sus vecinas (inferior y superior) a través de interfaces de capa a capa. En la figura 5 se muestra la comunicación física por medio de líneas continuas y la comunicación virtual mediante líneas discontinuas.

Nuestro interés principal en este libro apunta hacia la capa física del modelo OSI.

INTERNET

Toda exposición sobre redes de computadora lleva de manera natural a *Internet*. En el paradigma de Internet, la tecnología de red subyacente se desacopla de las aplicaciones corrientes al adoptar una definición abstracta de servicio de red. En términos más específicos, podemos establecer lo siguiente:

- Las aplicaciones se ejecutan independientemente de la tecnología empleada para construir la red.
- Mediante el mismo *token*, la tecnología de red es capaz de evolucionar sin afectar las aplicaciones.

La arquitectura de Internet, descrita en la figura 6, cuenta con tres bloques funcionales: computadoras hosts (anfitrionas), subredes y ruteadores. Las hosts constituyen nodos de la red, donde los datos se originan o donde se entregan. Los ruteadores conforman nodos intermedios que se emplean para cruzar las fronteras de las subredes. Dentro de una subred, todas las computadoras hosts que pertenecen a la subred intercambian datos directamente; observe, por ejemplo, las subredes 1 y 3 en la figura 6.

Al igual que otras redes de computadora, Internet tiene un conjunto de protocolos en capas. En particular, el intercambio de datos entre las hosts y los ruteadores se realiza por medio del *protocolo de Internet* (IP), como se ilustra en la figura 7. El IP es un protocolo universal que reside en la capa de la red (es decir, la capa 3 del modelo de referencia OSI); es simple y define un plan de direccionamiento con una capacidad integrada para transportar datos en la forma de paquetes de nodo a nodo. Al cruzar la frontera de la subred, los ruteadores toman las decisiones en cuanto a cómo deben designarse las rutas de los paquetes dirigidos a un destino específico. Esto se lleva a cabo con base en tablas de rutas que se desarrollan mediante el uso de protocolos a la medida para

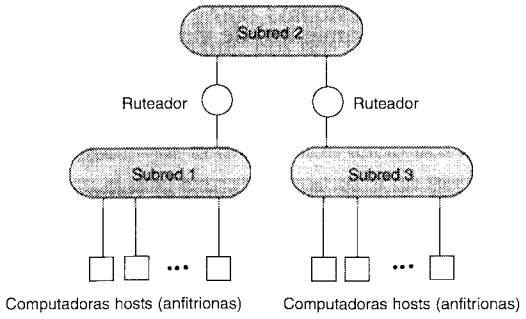


FIGURA 6 Una red interconectada de subredes.

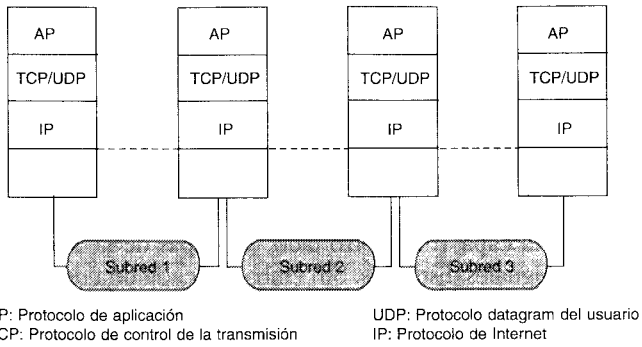


FIGURA 7 Ilustración de la arquitectura de red de Internet.

intercambiar información pertinente con otros ruteadores. El resultado neto del uso del conjunto de protocolos en capa es la entrega del *servicio del mejor esfuerzo*. Esto es, Internet ofrece entregar cada paquete de datos, pero no hay garantías en el tiempo de tránsito que se produzca en la entrega o incluso con respecto a si los paquetes se entregarán al receptor considerado.

■ REDES DE BANDA ANCHA

Con la creciente demanda de nuevos servicios (por ejemplo, video sobre demanda, comunicaciones multimedia) y la disponibilidad de tecnologías habilitadoras clave (entre ellas las fibras ópticas y los conmutadores digitales), la red telefónica evoluciona hacia una red de banda ancha para todos los fines que se conoce como *red digital de servicios integrados de banda ancha* (B-ISDN). La tecnología de soporte que hace posible a la B-ISDN es un protocolo de interfaz usuario-red denominado *modo de transferencia asíncrona* (ATM, por sus siglas en inglés). El ATM es una técnica de ancho de banda alto, retardo bajo y similar a los paquetes que se utiliza para conmutación y multiplexado; es independiente de los medios físicos de transporte. La característica de retardo bajo de la técnica se necesita para soportar servicios en tiempo real como la voz. La característica de ancho de banda alto se requiere para manejar video sobre demanda. En términos simples, el ATM es una tecnología oculta para los usuarios y un servicio orientado a la conexión que es visible para estos mismos.

Como su nombre lo indica, el ATM no es síncrono (es decir, en correspondencia con un reloj maestro). Posibilita el transporte de información digital en la forma de pequeños paquetes de tamaño fijo llamados *celdas*. El rasgo principal del ATM que debe destacarse aquí es que el servicio orientado a la conexión preserva el secuenciamiento de llamadas, lo que significa que no se necesita un reensamblado de celdas antes de presentar el flujo de tráfico a las hosts de destino. El despliegue de la tecnología de conmutación de celdas en la B-ISDN constituye un avance gigantesco en el uso tradicional de la conmutación de circuitos en la red telefónica.

El propósito fundamental del ATM es asignar *recursos de red* (es decir, ancho de banda, búfers y capacidad de procesamiento) de manera eficiente a fin de garantizar la *calidad de servicio* esperada en cada conexión. La calidad de servicio se mide en términos de tres parámetros:

- *Proporción de pérdida de celdas*, que se define como la proporción entre el número de celdas perdidas en el transporte a lo largo de la red y el número total de celdas bombeadas hacia la red.
- *Retardo de celda*, que se define como el tiempo que tarda una celda de una conexión particular en transitar a lo largo de la red.
- *Variación del retardo de celda*, que se define como la dispersión y/o inestabilidad en torno al retardo de celda medio.

La calidad de servicio que se ofrece en la B-ISDN se comparará con el *servicio de mejor esfuerzo* ofrecido por Internet.

TABLA 1 *Jerarquía de las velocidades de datos de la SONET*

Nivel[a]	*Velocidad de datos (Mb/s)*
OC-1	51.84
OC-3	155.52
OC-9	466.56
OC-12	622.08
OC-18	933.12
OC-24	1,244.16
OC-36	1,866.24
OC-48	2,488.32

[a] OC significa nivel de portadora óptica.

Luego de que se generan, las celdas ATM se estructuran para el transporte a lo largo de la red. Las celdas en la B-ISDN se ponen en un sistema de transmisión óptico denominado la *red óptica síncrona* (SONET);[8] las fibras ópticas se explican en la sección siguiente. La SONET utiliza multiplexado por división de tiempo, por lo que el ancho de banda completo de una fibra óptica se dedica a los diferentes flujos de datos entrantes en un esquema de tiempo compartido, de donde surge la necesidad de la operación síncrona. La SONET se controla por medio de un reloj maestro con una precisión de aproximadamente 1 parte en 10^9. De este modo, se envían bits de datos por una línea SONET a intervalos extremadamente precisos, que se controlan por medio de un reloj maestro. De cualquier modo, la SONET permite los arribos de tiempo irregular de las celdas del ATM.

La trama básica de la SONET es un bloque de 810 bytes que se lanza cada 125 μs para una velocidad de datos completa de 51.84 Mb/s. Tener 8000 tramas cada segundo corresponde exactamente a una velocidad de muestreo de 8 kHz, que es la velocidad de muestreo estándar para la transmisión digital de señales de voz por la red telefónica. Las velocidades de datos básicas de 51.84 Mb/s se intercalan síncronamente mediante bytes para generar una jerarquía de velocidades de datos, como se resume en la tabla 1.

5. *Canales de comunicación*

La transmisión de información por una red de comunicación se logra en la capa física por medio de un *canal de comunicación*. Dependiendo del modo de transmisión que se use, se distinguen dos grupos básicos de canales de comunicación: los que se basan en *propagación guiada* y los que se fundamentan en la *propagación libre*. El primer grupo incluye canales telefónicos, cables coaxiales y fibras ópticas. El segundo grupo comprende canales de transmisión inalámbrica, canales de radio móvil y canales de satélite. Los seis canales se describen a continuación.

i) Como se mencionó antes, una red telefónica ordinaria utiliza conmutación de circuitos para establecer un enlace de comunicación de extremo a extremo en una base temporal. El objetivo principal de la red es asegurar que la transmisión telefónica entre un hablante en un extremo del enlace y el escucha en el otro extremo constituya un sustituto aceptable para la conversación cara a cara. En esta forma de comunicación, la fuente del mensaje es el sonido que se produce mediante la voz del hablante y el destino final es el oído del escucha. No obstante, el canal telefónico soporta únicamente la transmisión de señales eléctricas. De acuerdo con lo anterior, se emplean los transductores apropiados en los extremos de transmisión y recepción del sistema. De manera específica, se coloca un *micrófono* cerca de la boca del hablante para

convertir las ondas sonoras en una señal eléctrica y esta señal se convierte de nuevo a la forma acústica mediante un *receptor de bobina móvil* situado cerca del oído del escucha. Los diseños actuales de estos dos transductores se han perfeccionado hasta responder bien a las frecuencias que varían de 20 a 8000 Hz; asimismo, es posible empaquetar en forma compacta un par de ellos dentro de un solo aparato telefónico con el que es fácil hablar o escuchar. El canal telefónico es uno de *ancho de banda limitado*. La restricción del ancho de banda surge del requerimiento de compartir el canal entre una multitud de usuarios al mismo tiempo. En consecuencia, una solución práctica al problema de la comunicación telefónica debe minimizar el requerimiento del ancho de banda del canal, sujeto a una transmisión satisfactoria de la voz humana. Para conseguirlo, las especificaciones de los transductores y del canal deben apegarse a estándares basados en pruebas subjetivas que se realizan en relación con la inteligibilidad o articulación de señales telefónicas mediante hablantes masculinos y femeninos representativos. Una señal de voz (masculina o femenina) está limitada en esencia a una banda de 300 a 3100 Hz a causa de que las frecuencias fuera de esta banda no contribuyen mucho a la eficiencia de la articulación. Por consiguiente, es posible considerar esta banda de frecuencia como una línea de guía aproximada para la pasobanda de un canal telefónico que proporciona un servicio satisfactorio, como se ilustra en la figura 8 para una conexión de larga distancia común. La figura 8*a* muestra las pérdidas de inserción del canal graficadas en función de la frecuencia; la *pérdida por inserción* (en dB) se define como $10 \log_{10}(P_0/P_L)$, donde P_L es la potencia entregada a una carga desde una fuente mediante el canal y P_0 es la potencia que se entrega a la misma carga cuando ésta se conecta directamente a la fuente. La figura 8*b* muestra la gráfica correspondiente del retardo (en milisegundos) de la envolvente (grupo) con respecto a la frecuencia; el *retardo de la envolvente* se define como el negativo de la derivada de la respuesta en fase con respecto a la frecuencia angular $\omega = 2\pi f$. Las gráficas de la figura 8 ilustran con claridad la naturaleza *dispersa* del canal telefónico.

El canal telefónico se construye utilizando pares trenzados para la transmisión de la señal. Un *par trenzado* está compuesto por dos conductores de cobre sólido, cada uno de los cuales se introduce en una funda de cloruro de polivinilo (PVC). Por lo general, cada par tiene una tasa de giro de 2 a 12 giros por pie, así como una impedancia característica de 90 a 110 ohms. Los pares trenzados suelen construirse dentro de cables, con cada cable compuesto de muchos pares muy próximos unos de otros. Los pares trenzados son natural-

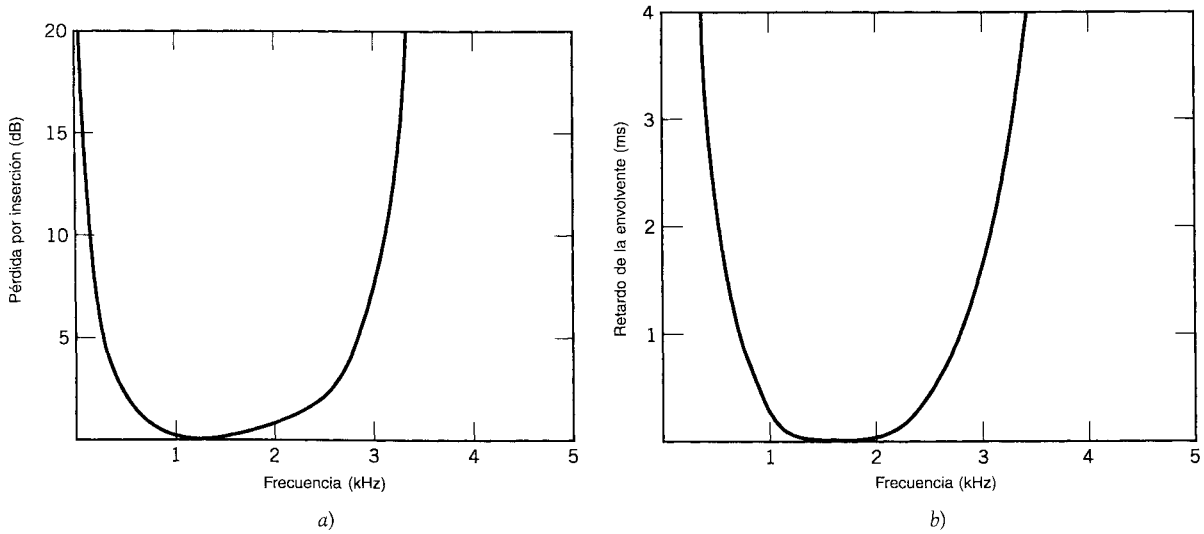


FIGURA 8 Características de la conexión telefónica ordinaria: *a*) pérdida por inserción, *b*) retardo de la envolvente. (Adaptada de Bellamy.)

mente susceptibles a la *interferencia electromagnética* (EMI), cuyos efectos se mitigan mediante el trenzado de los alambres.

ii) Un *cable coaxial* consta de un conductor interior y de un conductor exterior, separados por un material aislante dieléctrico. El conductor interior está hecho de alambre de cobre que se introduce en el material dieléctrico. El conductor exterior es de cobre estañado o de acero revestido de cobre. Por ejemplo, un cable coaxial tiene una impedancia característica de 50 o 75 ohms. Comparado con un cable de par trenzado, el cable coaxial ofrece un mayor grado de inmunidad a la interferencia electromagnética. Además, en virtud de su mucho más alto ancho de banda, los cables coaxiales pueden soportar la transmisión de datos digitales a velocidades de bit mucho más elevadas que los pares trenzados. Resultan factibles velocidades de hasta 20 Mb/s utilizando cables coaxiales, siendo 10 Mb/s el estándar.

En tanto que el uso del par trenzado se ha confinado principalmente al servicio punto a punto, el cable coaxial tiene la posibilidad de operar como un medio de acceso múltiple utilizando derivaciones de alta impedancia. Una aplicación común de los cables coaxiales es como medio de transmisión para las redes de área local en un ambiente de oficinas.

Otra aplicación común de los cables coaxiales corresponde a los *sistemas de televisión por cable*, conocidos también como *sistemas de televisión de antena comunitaria* (CATV). En esta aplicación los cables coaxiales se emplean para distribuir las señales de televisión, audio y datos desde el extremo principal hasta los suscriptores. El *extremo principal* es la unidad de origen central del sistema CATV, donde se transportan y procesan todas las señales.

iii) Una *fibra óptica* es una guía de onda dieléctrica que transporta señales luminosas desde un lugar a otro del mismo modo que un cable de par trenzado o coaxial transporta señales eléctricas. Está compuesto por un *núcleo central* dentro del cual el campo electromagnético que se propaga queda confinado y está rodeado por una capa de *revestimiento metálico*, rodeada a su vez por un delgado *forro de protección*.[9] El núcleo y el revestimiento metálico se elaboran en ambos casos a partir de vidrio de sílice, en tanto que el forro es de plástico. Las fibras ópticas cuentan con características únicas que las hacen altamente atractivas como medio de transmisión. En particular, ofrecen las siguientes características únicas:

- *Ancho de banda de enorme potencial*, que resulta del uso de frecuencias de portadora óptica de alrededor de 2×10^{14} Hz; con una frecuencia de portadora tan alta y un ancho de banda aproximadamente igual a 10% de la frecuencia de la portadora, el ancho de banda teórico del sistema de ondas luminosas es cercano a 2×10^{13} Hz, que es realmente muy grande.
- *Pérdidas de transmisión bajas*, tan bajas como 0.1 dB/km.
- *Inmunidad a la interferencia electromagnética*, que es una característica inherente de una fibra óptica vista como una guía de onda dieléctrica.
- *Tamaño y peso pequeños*, lo que está caracterizado por un diámetro que no es mayor al de un cabello humano.
- *Resistencia y flexibilidad*, lo que se ejemplifica mediante resistencias a la tensión muy alta y la posibilidad de flexionarlas o torcerlas sin que se produzca daño.

Un último aspecto, aunque no por eso menos importante, se refiere al hecho de que las fibras ópticas ofrecen el potencial para comunicaciones con líneas de bajo costo, ya que éstas se fabrican a partir de arena, la que, a diferencia del cobre que se utiliza en los conductores metálicos, no es un recurso que escasee. Las propiedades únicas de las fibras ópticas han impulsado enormes avances en la tecnología de sistemas de ondas luminosas, los cuales han revolucionado las comunicaciones a larga distancia y continúan en ese camino.

iv) Los *canales de transmisión inalámbricos* sustentan la transmisión de señales de radio y televisión. La señal que transporta la información, lo que representa el habla, la música o las imágenes, se modula sobre una frecuencia de portadora que identifica a la estación transmisora; la modulación se describe en la sección siguiente. La transmisión se origina a partir de una *antena* que actúa como la unidad de transición o acoplamiento entre la fuente de la señal modulada y las ondas electromagnéticas en el espacio libre. El objetivo

al diseñar la antena es excitar las ondas en la dirección o direcciones requeridas, del modo más eficiente posible. Por lo general la antena transmisora se monta sobre una torre para proporcionar una vista sin obstáculos del área de los alrededores, lo más lejos posible. Debido al fenómeno de la *difracción*, que es una propiedad fundamental del movimiento ondulatorio, las ondas de radio se desvían en torno a la superficie terrestre. La propagación más allá de la línea de vista se hace por ello posible, aunque con un poco más de pérdidas que las que se producen en el espacio libre.

En el extremo de recepción se emplea una antena para captar las ondas radiadas, estableciéndose un enlace de comunicación con el transmisor. La mayor parte de los receptores de radio son del tipo *superheterodino*. Esta técnica consiste en convertir la señal recibida dentro de cierta banda de frecuencias conveniente, denominada la *banda de frecuencia intermedia* (IF, por sus siglas en inglés), y en recuperar después la señal original que transporta la información por medio de un detector apropiado.

v) Un *canal de radio móvil* amplía la capacidad de la red de telecomunicaciones pública introduciendo *movilidad* en la red debida a su capacidad para transmitir. El término *radio móvil* normalmente significa abarcar situaciones terrestres donde un transmisor o receptor de radio es capaz de moverse, independientemente de que se desplace o no. Los principales efectos de propagación que se encuentran en el uso de un radio móvil en áreas pobladas se deben a que la antena de la unidad móvil quizá se encuentre bastante debajo de las edificaciones de los alrededores. En términos simples, no hay una trayectoria de "línea de vista" para la comunicación; en vez de eso, la propagación de radio ocurre principalmente mediante la dispersión en las superficies de las construcciones de los alrededores y a partir de la difracción por y/o alrededor de ellas. El resultado final es que la energía llega a la antena receptora por medio de más de una trayectoria. En un ambiente de radio móvil, hablamos consecuentemente de un *fenómeno de trayectorias múltiples* en el que las diversas ondas de radio entrantes llegan a su destino desde diferentes direcciones y con diferentes tiempos de retardo. En realidad, es posible que exista una multitud de trayectorias de propagación con diferentes longitudes eléctricas, y sus contribuciones a la señal recibida podrían combinarse en una diversidad de formas. En consecuencia, la intensidad de la señal recibida varía con la ubicación de una manera complicada, y es por eso que un canal de radio móvil se considera como uno *lineal variante con el tiempo* que es de naturaleza estadística.

vi) Por último, un *canal de satélite* añade otra dimensión invaluable a la red de telecomunicaciones pública al proporcionar una *cobertura de área amplia* tanto en el sentido continental como en el intercontinental. Asimismo, el acceso a áreas remotas no cubiertas mediante comunicaciones de cable convencional o de fibra es también un rasgo distintivo de los satélites. En casi todos los sistemas de comunicación de satélite, los satélites se sitúan en *órbita geoestacionaria*. Se necesita satisfacer dos requerimientos para que la órbita sea de este tipo. Primero, la órbita es *geosíncrona*, lo que requiere que el satélite se encuentre a una altura de 22,300 millas; un satélite geosíncrono orbita la Tierra en 24 horas (es decir, el satélite está en sincronía con la rotación terrestre). Segundo, el satélite se coloca en una órbita directamente arriba del ecuador apuntando hacia el este (es decir, con inclinación cero). Visto desde la Tierra, un satélite en órbita geoestacionaria parece estar fijo en el cielo. En consecuencia, una estación terrestre *no* tiene que seguir al satélite; en vez de eso, sólo tiene que apuntar su antena a lo largo de una dirección fija, directamente hacia el satélite. Al hacerlo de esa manera, el diseño del sistema se simplifica de modo considerable. Los satélites de comunicaciones en órbita geoestacionaria ofrecen las siguientes capacidades únicas del sistema:

- Cobertura de área amplia.
- Enlaces de transmisión confiables.
- Anchos de banda de transmisión amplios.

En términos de servicios, los satélites tienen la posibilidad de ofrecer enlaces fijos de punto a punto que se extienden por largas distancias y hasta áreas remotas, comunicación a plataformas móviles (por ejemplo, aviones, barcos) o capacidades de transmisión. En realidad, los satélites de comunicación desempeñan un papel fundamental en la noción del mundo completo que se está considerando como una "aldea global". En un sistema de comunicación de satélite característico, una señal de mensaje se transmite desde una estación terrestre por medio de un *enlace ascendente* hasta un satélite, se amplifica en un *transpondedor* (es decir, circuitería electrónica) a bordo del satélite y luego se retransmite desde el satélite por medio de un *enlace descendente* hasta otra estación terrestre. Con el satélite situado en una órbita geoestacionaria, siempre son visibles todas las estaciones terrestres ubicadas dentro de las zonas de cobertura de la antena del satélite sobre la superficie de la Tierra. En efecto, el satélite actúa como una poderosa *repetidora* en el cielo. La banda de frecuencias más popular para las comunicaciones de satélite es la de 6 GHz para el enlace ascendente y de 4 GHz para el descendente. El uso de esta banda de frecuencia ofrece los siguientes atributos:

- Equipo de microondas relativamente económico.
- Atenuación baja debido a la lluvia; esta última es una causa atmosférica principal de la pérdida de señal.
- Ruido de fondo del cielo insignificante; el ruido de fondo del cielo (debido a las emisiones de ruido desde fuentes galácticas, del sol y terrestres) alcanza su nivel más bajo entre 1 y 10 GHz.

En la banda de 6/4 GHz, a un satélite representativo se le asigna un ancho de banda de 500 MHz que se divide entre los 12 transpondedores a bordo del satélite. Cada uno de estos últimos, utilizando alrededor de 36 MHz del ancho de banda del satélite, corresponde a un canal de radio específico. Es posible que un solo transpondedor transporte al menos una señal de televisión a color, 1200 circuitos de voz o datos digitales a una velocidad de 50 Mb/s.

En resumen, un canal de comunicación es fundamental para la operación de un sistema de comunicaciones. Sus propiedades determinan tanto la capacidad de transporte de información del sistema como la calidad de servicio que ofrece este mismo. Es factible clasificar los canales de comunicación de diferentes maneras:

- Un canal puede ser *lineal* o *no lineal*; un canal de radio inalámbrico es lineal, en tanto que uno de satélite suele ser (aunque no siempre) no lineal.
- Es posible que un canal sea *invariante con el tiempo* o *variante con el tiempo*; una fibra óptica es invariante con el tiempo en tanto que un canal de radio móvil suele ser por lo general variante con el tiempo.
- Un canal quizá sea de *ancho de banda limitado* o de *potencia limitada* (es decir, limitado en la potencia transmitida disponible); un canal telefónico es de ancho de banda limitado, mientras que un canal de enlace de fibra óptica y uno de satélite son en ambos casos de potencia limitada.

Luego de que disponemos de cierta comprensión de las fuentes de información y de los canales de comunicación, es factible regresar al diagrama de bloques de un sistema de comunicaciones que se muestra en la figura 1.

6. *Proceso de modulación*

El fin de un sistema de comunicaciones es entregar una señal de mensaje desde una fuente de información en una forma reconocible hasta el destino del usuario, con la fuente y el usuario separados físicamente uno del otro. Para realizar lo anterior, el transmisor modifica la señal del mensaje en una forma adecuada para la transmisión por el canal. Esta modificación se logra mediante un proceso conocido como *modulación*, que implica variar algún parámetro de la *onda portadora* de acuerdo con la señal del mensaje. El receptor recrea la señal del mensaje original a partir de una versión degrada-

da de la señal transmitida luego de la propagación a través del canal. Esta recreación se consigue al emplear un proceso conocido como *demodulación*, que es el inverso del proceso de modulación utilizado en el transmisor. Sin embargo, debido a la inevitable presencia de ruido y de distorsión en la señal recibida, encontramos que el receptor no puede recrear exactamente la señal del mensaje original. La degradación que resulta en el desempeño completo del sistema depende del tipo del esquema de modulación que se utilice. De modo específico, observamos que algunos esquemas de modulación son menos sensibles a los efectos del ruido y la distorsión que otros.

Es posible clasificar el proceso de modulación en *modulación de onda continua* y *modulación por pulsos*. En la modulación de onda continua (CW) se emplea una onda senoidal como la portadora. Cuando la amplitud de la portadora varía de acuerdo con la señal del mensaje, tenemos *modulación de amplitud* (AM), y cuando se varía el ángulo de la portadora, tenemos *modulación angular*. Es posible subdividir aún más la última forma de modulación de onda continua en *modulación de frecuencia* (FM) y *modulación de fase* (PM), en las cuales la frecuencia y fase instantáneas de la portadora, respectivamente, se varían de acuerdo con la señal del mensaje.

Por otro lado, en la modulación por pulsos la portadora está compuesta por una secuencia periódica de pulsos rectangulares. La modulación por pulsos puede ser por sí sola de tipo analógico o digital. En la *modulación por pulsos analógica*, la amplitud, duración o posición de un pulso varía de acuerdo con valores de muestra de la señal del mensaje. En un caso de este tipo, hablamos de la *modulación por amplitud de pulsos* (PAM), *modulación por duración de pulsos* (PDM) y *modulación por posición de pulsos* (PPM).

La forma digital estándar de la modulación por pulsos se conoce como *modulación por codificación de pulsos* (PCM) que no tiene contraparte de onda continua. La PCM se inicia esencialmente como la PAM, pero con una modificación importante. La amplitud de cada pulso modulado (es decir, la *muestra* de la señal del mensaje original) se *cuantiza* o *redondea* hasta el valor más cercano en un conjunto preestablecido de niveles de amplitud *discretos* y luego se *codifica* en una secuencia correspondiente de símbolos binarios. Los propios símbolos binarios 0 y 1 se representan mediante señales de pulsos a los que se les da una forma adecuada para la transmisión por el canal. En cualquier acontecimiento, como consecuencia del proceso de cuantización, siempre se pierde cierta información y la señal del mensaje original no puede reconstruirse con exactitud. A pesar de eso, siempre que el número de niveles de cuantización (amplitud discreta) es suficientemente grande, la distorsión que produce el proceso de cuantización no la distingue el oído humano en el caso de una señal de voz o el ojo humano en el caso de una imagen bidimensional. Entre todos los esquemas de modulación diferentes, la decodificación por pulsos ha surgido como el método de modulación preferido para la transmisión de señales de mensaje analógicas por las siguientes razones:

- *Robustez* en ambientes ruidosos mediante la *regeneración* a intervalos regulares de la señal transmitida.
- Operación *flexible*.
- *Integración* de diversas fuentes de información en un formato común.
- *Seguridad* de la información y de su transmisión desde la fuente hasta el destino.

Al introducir la idea de modulación, subrayamos su importancia como un proceso que asegura la transmisión de la señal del mensaje por un canal prescrito. Existe otro beneficio importante, a saber, el *multiplexado*, que resulta del empleo de la modulación. El multiplexado es el proceso de combinar varias señales del mensaje para su transmisión simultánea por el mismo canal. Los siguientes son tres métodos de multiplexado que se usan comúnmente:

- *Multiplexado por división de frecuencia* (FDM): la modulación de onda continua se utiliza para trasladar cada señal del mensaje a fin de que resida en una ranura de frecuencia específica dentro de la pasobanda del canal asignándole una frecuencia de portadora dis-

tinta. En el receptor se recurre a un banco de filtros para separar las diferentes señales moduladas y prepararlas individualmente para la demodulación.

- *Multiplexado por división de tiempo* (TDM): la modulación por pulsos se emplea para colocar muestras de señales de mensajes diferentes en ranuras de tiempo que no se traslapan.
- *Multiplexado por división de código* (CDM): cada señal del mensaje se identifica mediante un código distintivo.

En el FDM las señales del mensaje se traslapan entre sí a la entrada del canal. Por tanto, es posible que el sistema sufra de *diafonía* (es decir, la interacción entre señales de mensaje) si el canal es no lineal. En el TDM las señales de mensaje utilizan la pasobanda completa del canal, pero en un esquema de tiempo compartido. En el CDM se permite que las señales de mensaje se traslapen tanto en tiempo como en frecuencia a lo largo del canal.

Debe mencionarse el *multiplexado por división de longitud de onda* (WDM), que es especial para las fibras ópticas. En este tipo de multiplexado, la longitud de onda se utiliza como un nuevo grado de libertad al operar en forma concurrente distintas porciones del espectro de longitudes de onda (es decir, *colores* distintos) a los que se tiene acceso dentro de la fibra óptica. Sin embargo, al reconocer la relación recíproca que existe entre la longitud de onda y la frecuencia de una onda electromagnética, es factible afirmar que el WDM es una forma de FDM.

7. *Tipos de comunicación analógica y digital*

En la mayoría de los casos, en el diseño de un sistema de comunicación, la fuente de información, el canal de comunicación y el receptor de información (usuario final) están especificados por completo. El reto es diseñar el transmisor y el receptor con las siguientes líneas guía en mente:

- Codificar/modular la señal de mensaje generada por la fuente de información, transmitirla por el canal y producir una "estimación" de ella en la salida del receptor que satisfaga los requerimientos del usuario final.
- Efectuar todo lo anterior a un costo viable.

Tenemos la opción de utilizar un sistema de comunicación digital o analógico.

Considere primero el caso de un sistema de *comunicación digital* representado por el diagrama de bloques de la figura 9, cuya explicación está inmersa en la teoría de la información. Los bloques funcionales del transmisor y el receptor, empezando desde el extremo lejano del canal, se vinculan del modo siguiente:

- *Codificador-decodificador de fuente.*
- *Codificador-decodificador de canal.*
- *Modulador-demodulador.*

El codificador de fuente elimina información redundante de la señal de mensaje y es responsable del uso eficiente del canal. La secuencia de símbolos resultante se denomina *la palabra de código fuente*. El codificador del canal procesa posteriormente el flujo de datos, lo cual produce una nueva secuencia de símbolos denominada la *palabra de código de canal*. Ésta es más grande que la palabra de código fuente en virtud de la redundancia *controlada* integrada en su construcción. Por último, el modulador representa cada símbolo de la palabra de código de canal mediante un símbolo analógico correspondiente, elegido de manera apropiada a partir de un conjunto finito de símbolos analógicos posibles. La secuencia de símbolos analógicos producidos por el modulador se denomina *forma de onda*, la cual es adecuada para la transmisión por el canal. En el receptor, la salida del canal (señal recibida) se procesa en orden inverso al del transmisor, reconstruyendo de esa manera una versión reconocible de la señal de mensaje original. La señal de mensaje reconstruida se entrega finalmente al usuario de información en el

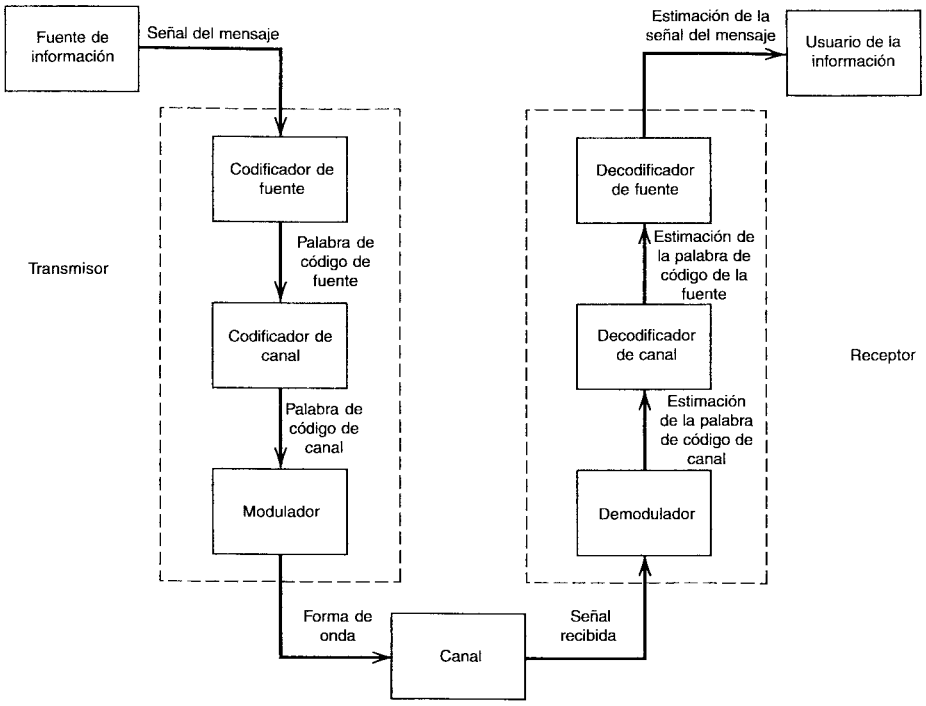


FIGURA 9 Diagrama de bloques de un sistema de comunicación digital.

destino. A partir de esta descripción queda claro que el diseño de un sistema de comunicación digital es bastante complejo en los términos conceptuales, pero fácil de construir. Además, el sistema es *robusto*, lo que ofrece una mayor tolerancia a los efectos físicos (por ejemplo, variaciones de temperatura, envejecimiento, vibraciones mecánicas) que su contraparte analógica.

En contraste, el diseño de un *sistema de comunicación analógico* es simple en términos conceptuales, aunque difícil de construir debido a los estrictos requerimientos en la linealidad y el ajuste del sistema. Por ejemplo, la comunicación de voz requiere productos con distorsión no lineal al menos a 40 dB debajo de la señal de mensaje deseada. En términos de procesamiento de la señal, el transmisor consiste en un modulador y el receptor en un demodulador, cuyos detalles se determinan mediante el tipo de modulación de onda continua utilizado.

La simplicidad conceptual de las comunicaciones analógicas se debe a que las técnicas de modulación analógica, ejemplificadas por su extenso uso en el radio y la televisión, hacen relativamente superficiales los cambios a la señal de mensaje a fin de prepararla para la transmisión por el canal. De modo más específico, el diseñador del sistema no requiere de un esfuerzo importante para ajustar la forma de onda de la señal transmitida para adaptar el canal a cualquier nivel más profundo. Por otra parte, la teoría de la comunicación digital intenta determinar un conjunto finito de formas de onda que correspondan estrechamente con las características del canal y que, en consecuencia, son más tolerantes a los deterioros de este último. Al hacerlo de esa manera, se establece una comunicación confiable por el canal. En la selección de buenas formas de onda para comunicación digital por un canal ruidoso, el diseño se ve afectado únicamente por las características del canal. Sin embargo, una vez que se ha elegido el conjunto apropiado de formas de onda para transmisión por el canal, es posible codificar la información de la fuente en las formas de onda del canal y de ese modo asegurar la transmisión eficiente de

información desde la fuente hasta el usuario. En resumen, el uso de comunicaciones digitales ofrece la capacidad para la transmisión de información que es tanto *eficiente* como *confiable*.

A partir de esta exposición, resulta patente que el uso de las comunicaciones digitales requiere una cantidad considerable de circuitería electrónica, aunque la electrónica de la actualidad es económica, debido a la cada vez mayor disponibilidad de circuitos integrados a muy grande escala de integración (VLSI) en la forma de chips de silicio. Así, a pesar de que en el pasado las consideraciones de costo solían ser un factor en la selección de comunicaciones analógicas con respecto a las comunicaciones digitales, éste ya no es el caso.

A pesar de la tendencia hacia el uso siempre creciente de las comunicaciones digitales, es posible considerar un fuerte argumento para estudiar las comunicaciones analógicas por dos razones importantes:

1. Mientras escuchamos y vemos comunicaciones analógicas a nuestro alrededor por medio de la radio y la televisión, necesitamos comprender cómo funcionan estos sistemas de comunicación. Además, el estudio de la modulación analógica estimula otros esquemas de modulación digital.
2. Los dispositivos y los circuitos analógicos tienen una afinidad natural para operar a muy altas velocidades y consumen muy poca potencia en comparación con sus contrapartes digitales. Por consiguiente, la puesta en práctica de sistemas de comunicación de muy alta velocidad o muy baja potencia impone el uso del método analógico.

8. *Teorema de la capacidad de información de Shannon*

El objetivo de un diseñador de sistemas de comunicación es configurar un sistema que transporte una señal de mensaje de una fuente de interés a través de un canal ruidoso hasta un usuario en el otro extremo del canal con el siguiente objetivo:

> La señal de mensaje se entrega al usuario tanto eficiente como confiablemente, con apego a ciertas restricciones de diseño: potencia de transmisión permisible, ancho de banda del canal disponible y costo viable de construcción del sistema.

En el caso de un sistema de comunicación digital, la confiabilidad se expresa comúnmente en términos de la *tasa de error de bits* (BER) o *probabilidad de error de bit* medida en la salida del receptor. Claramente, cuanto más pequeña es la BER, tanto más confiable resulta el sistema de comunicación. Una pregunta que viene a la mente en este contexto es si es posible diseñar un sistema de comunicación que opere con BER cero incluso a través de una canal ruidoso. En un escenario *ideal*, la respuesta a esta pregunta es un *sí* enfático. La respuesta está embebida en uno de los célebres teoremas de Shanon,[10] que recibe el nombre de teorema de la capacidad de información.

Dejemos que B denote el ancho de banda del canal y que SNR indique la relación señal a ruido recibida. El *teorema de capacidad de información* establece que estos dos parámetros se relacionan idealmente como

$$C = B \log_2(1 + \text{SNR}) \ b/s \tag{1}$$

donde C es la capacidad de información del canal. La *capacidad de información* se define como la velocidad máxima a la cual puede transmitirse sin error la información a lo largo del canal; se mide en *bits por segundo* (b/s). Para un ancho de banda de canal preestablecido B y una SNR recibida, el teorema de capacidad de información nos indica que una señal de mensaje puede transmitirse a través del sistema sin error aun cuando el canal sea ruidoso, siempre que la *velocidad de señalización real* R, en bits por segundo, a la cual se transmiten los datos por el canal, sea menor que la capacidad de información C.

Desafortunadamente, el teorema de capacidad de información de Shannon no nos dice cómo diseñar el sistema. A pesar de eso, desde un punto de vista de diseño, el teorema resulta muy valioso por las siguientes razones:

1. El teorema de capacidad de información proporciona una cota con respecto a qué velocidad de transmisión de datos es teóricamente alcanzable para valores preestablecidos del ancho de banda del canal B y de la SNR recibida. Con base en ello, es posible utilizar la proporción
$$\eta = \frac{R}{C}$$
como una medida de la *eficiencia* del sistema de comunicación digital en estudio. Cuanto más cercana es η a la unidad, tanto más eficiente es el sistema.
2. La ecuación (1) ofrece una base para el compromiso entre el ancho de banda del canal B y la SNR recibida. En particular, para una velocidad de señalización preestablecida R, resulta viable reducir la SNR requerida aumentando el ancho de banda del canal B, de lo cual surge la motivación para utilizar un esquema modulado de banda ancha (por ejemplo, modulación por codificación de pulsos) para mejorar el desempeño relativo al ruido.
3. La ecuación (1) proporciona un marco de referencia idealizado para comparar el desempeño relativo al ruido de un esquema de modulación con otro.

9. *Un problema de comunicación digital*

Cuando hablamos de un sistema de comunicación digital que tiene una tasa de error de bits baja, por ejemplo, la implicación es que el receptor decodifica erróneamente sólo una pequeña fracción en un largo flujo de símbolos binarios. El aspecto del receptor que determina si un símbolo binario enviado por un canal ruidoso se decodifica en error o no es de importancia fundamental para el diseño de sistemas de comunicación digitales. En consecuencia, resulta apropiado explicar brevemente este aspecto básico al grado de motivar el estudio de los sistemas de comunicación.

Suponga que tenemos una señal binaria aleatoria, $m(t)$, consistente en símbolos 1 y 0 que son igualmente probables. El símbolo 1 se representa mediante un nivel constante $+1$, y el símbolo 0 se representa mediante un nivel constante -1, cada uno de los cuales tiene una duración T. Una señal de estas características quizá represente la salida de una computadora digital o la versión digitalizada de una señal de voz. Para facilitar la transmisión de esta señal por un canal de comunicación, empleamos un esquema de modulación simple conocido como *corrimiento de fase por llaveo*. Específicamente, la información que lleva la señal $m(t)$ se multiplica por una onda portadora senoidal $A_c \cos(2\pi f_c t)$, donde A_c es la amplitud de la portadora, f_c es la frecuencia de la portadora y t corresponde al tiempo. La figura 10*a* muestra un diagrama de bloques del transmisor, la salida del cual se define por medio de

$$s(t) = \begin{cases} A_c \cos(2\pi f_c t) & \text{para el símbolo 1} \\ -A_c \cos(2\pi f_c t) & \text{para el símbolo 0} \end{cases} \tag{2}$$

donde $0 \le t \le T$. La frecuencia de la portadora f_c es un múltiplo de $1/T$.

Se supone que el canal no tiene distorsión, aunque es ruidoso, según se describe en la figura 10*b*. La señal recibida $x(t)$ está definida consecuentemente por

$$x(t) = s(t) + w(t) \tag{3}$$

donde $w(t)$ es el ruido de canal aditivo.

El receptor consta de un correlator seguido por un dispositivo de toma de decisiones, como se indica en la figura 10*c*. El *correlator* multiplica la señal recibida $x(t)$ por una portado-

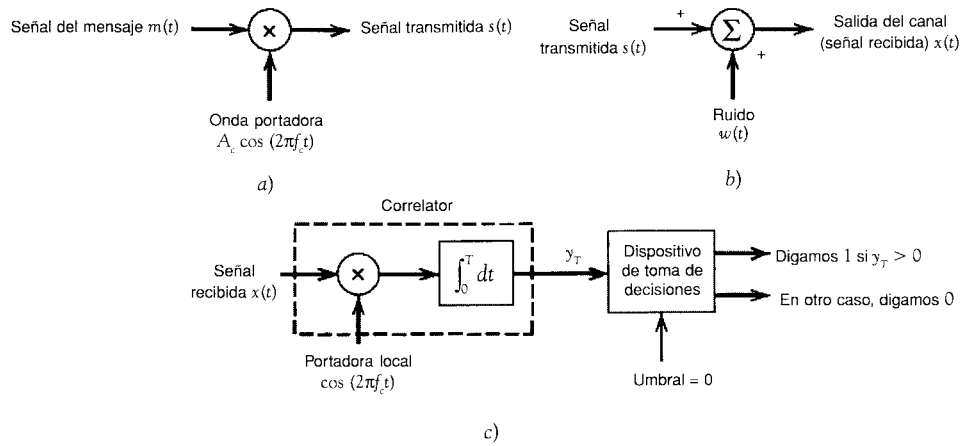


FIGURA 10 Elementos de un sistema de comunicación digital. *a*) Diagrama de bloques de transmisor. *b*) Diagrama de bloques de canal. *c*) Diagrama de bloques de receptor.

ra $\cos(2\pi f_c t)$ generada localmente e integra después el producto sobre el intervalo del símbolo $0 \le t \le T$, produciendo la salida

$$y_T = \int_0^T x(t) \cos(2\pi f_c t)\, dt \tag{4}$$

Al sustituir las ecuaciones (2) y (3) en la (4) y bajo la suposición de que la frecuencia de la portadora f_c es un múltiplo de $1/T$, obtenemos (después de la simplificación de términos)

$$y_T = \begin{cases} +\dfrac{A_c}{2} + w_T & \text{para el símbolo 1} \\ -\dfrac{A_c}{2} + w_T & \text{para el símbolo 0} \end{cases} \tag{5}$$

donde w_T es la contribución de la salida del correlator debida al ruido del canal $w(t)$. Para reconstruir la señal binaria original $m(t)$, la salida del correlator y_T se compara con un umbral de cero volts mediante el *dispositivo de toma de decisiones*, cuya operación se basa en la siguiente regla:

> Si la salida de y_T es mayor que cero, el receptor produce la salida símbolo 1; en otro caso, la salida es el símbolo 0.

Con este antecedente, es posible discutir o plantear ahora algunos aspectos básicos. En primer lugar, a partir del análisis de Fourier encontramos que el *producto de tiempo-ancho de banda* de una señal de pulsos es constante. Esto significa que el ancho de banda de un pulso rectangular de duración T es inversamente proporcional a T. La señal transmitida en la figura 10*a* consiste en el producto de esta señal rectangular y de la portadora senoidal $A_c \cos(2\pi f_c t)$. La multiplicación de una señal por una senoide tiene el efecto de correr a la derecha la transformada de Fourier de la señal en f_c y a la izquierda por una cantidad igual, salvo por el factor de escalamiento de 1/2. Se concluye, en consecuencia, que el ancho de banda de la señal transmitida $m(t)$, y por lo tanto el ancho de banda del canal requerido, es inversamente proporcional al recíproco de la duración del símbolo T. Para el problema presente, el recíproco de T es también la velocidad de señalización del sistema en b/s.

No obstante, existen algunos otros aspectos que requieren consideraciones teóricas:

1. ¿Cuál es la justificación para la estructura del receptor de la figura 10c?
2. La contribución del ruido w_T es el valor de una variable aleatoria W producida muestreando una cierta realización $w(t)$ del ruido del canal en el tiempo $t = T$ de acuerdo con las ecuaciones (3) y (4). ¿Cómo relacionar las estadísticas de la variable aleatoria W con las características estadísticas del ruido del canal?
3. El receptor de la figura 10c produce errores ocasionales debido a la naturaleza aleatoria de la salida del correlator. Esto es, el receptor decide en favor del símbolo 0 dado que el símbolo 1 se transmitió realmente y viceversa. ¿Cuál es la probabilidad de los errores de decisión?

Asimismo, existen algunos aspectos prácticos importantes que necesitan atención:

1. El ancho de banda del canal es un recurso altamente valioso. ¿Cómo elegimos un esquema de modulación que conserve el ancho de banda a un costo adecuado?
2. La señal binaria $m(t)$ quizá incluya símbolos redundantes introducidos en ella mediante el uso de la codificación del canal al grado de proporcionar protección contra el ruido del canal. ¿Cómo diseñamos el codificador del canal en el transmisor y el decodificador del canal en el receptor de tal modo que nos acerquemos al teorema de la capacidad de información de Shannon para que sea físicamente realizable?
3. La portadora generada en forma local en el receptor de la figura 10c está separada físicamente de la fuente de la portadora que se utiliza para la modulación en el transmisor. ¿Cómo *sincronizamos* el receptor con el transmisor en lo que respecta tanto a la fase portadora como a la temporización del símbolo para que se justifique el uso de la ecuación (4) como la base de la toma de decisiones en la reconstrucción de la señal binaria original?

Los aspectos teóricos y prácticos que se han planteado aquí en el contexto del sistema de comunicación digital simple de la figura 10 se consideran en los siguientes capítulos del libro.

10. *Notas históricas*[11]

Una presentación preliminar de las comunicaciones sería incompleta sin una historia del tema. En la sección final de este capítulo introductorio presentamos algunas notas históricas acerca de las comunicaciones; cada párrafo se enfoca en algunos acontecimientos relacionados importantes. Se espera que este material proporcione una sensación de inspiración y estímulo para el lector.

En 1837, Samuel Morse, un pintor, perfeccionó el telégrafo. Con las palabras "Qué ha forjado dios", transmitidas por el telégrafo eléctrico de Morse entre Washington, D.C., y Baltimore, Maryland, en 1844, surgieron medios por completo revolucionarios de comunicaciones en tiempo real a larga distancia. El telégrafo es el precursor de las comunicaciones digitales en cuanto a que el *código Morse* es ternario de *longitud variable* y utiliza un alfabeto de cuatro símbolos: un punto, una raya, un espacio de letra y un espacio de palabra; las secuencias cortas representan letras frecuentes, en tanto que las largas corresponden a letras poco frecuentes. Este tipo de transmisión de señales es ideal para el tecleo manual. Subsecuentemente, Emile Baudot creó un código binario de *longitud fija* para el telégrafo en 1875. En el código telegráfico de Baudot, bastante adecuado para teletipos, cada palabra de código consta de cinco elementos de código de longitud igual y a cada elemento se le asigna uno de dos estados posibles: una *marca* o un *espacio* (es decir, símbolo 1 o 0 en la terminología actual).

En 1864, James Clerk Maxwell formuló la teoría electromagnética de la luz y predijo la existencia de ondas de radio; el conjunto fundamental de ecuaciones lleva su nombre. La existencia de ondas de radio la estableció experimentalmente Heinrich Hertz en 1887. En 1894,

Oliver Lodge demostró la comunicación inalámbrica a lo largo de una distancia relativamente corta (150 yardas). Posteriormente, el 12 de diciembre de 1901, Guglielmo Marconi recibió una señal de radio en Signal Hill en Terranova; la señal de radio se había originado en Cornwall, Inglaterra, a 1700 millas de distancia a través del Atlántico. Se había abierto de ese modo el camino hacia una tremenda ampliación del alcance de las comunicaciones. En 1906, Reginald Fessenden, un académico autodidacta, hizo historia al conducir la primera transmisión de radio.

En 1875, Alejandro Graham Bell, un maestro de sordos, inventó el *teléfono*. Este aparato convirtió en una realidad práctica la transmisión en tiempo real del habla mediante la codificación eléctrica y la replicación del sonido. La primera versión del teléfono fue burda y limitada, y únicamente posibilitaba que la gente hablara a cortas distancias. Cuando el servicio telefónico contaba apenas con unos cuantos años, se generó el interés por automatizarlo. Notablemente, en 1897, A.B. Strowger, contratista de Kansas City, Missouri, ideó el *conmutador por grados* automático que lleva su nombre; de todos los conmutadores electromecánicos ideados a lo largo de los años, el conmutador Strowger fue el más popular y el que se utilizó más ampliamente.

En 1904, John Ambrose Fleming inventó el *diodo de tubo de vacío*, que abrió paso a la invención del *triodo de tubo de vacío* de Lee de Forest en 1906. El descubrimiento del triodo marcó los albores de las comunicaciones de voz inalámbricas. En realidad, hasta la invención y perfección del transistor, el triodo fue el dispositivo supremo en el diseño de los amplificadores electrónicos.

En 1918, Edwin H. Armstrong inventó el receptor de *radio superheterodino*; hasta ahora casi todos los receptores de radio son de este tipo. En 1933, Armstrong demostró otro concepto revolucionario, a saber, un esquema de modulación que denominó *modulación de frecuencia* (FM); el artículo de Armstrong que fundamentó el radio FM se publicó en 1936.

Philo T. Farnsworth probó el primer sistema de *televisión* totalmente electrónico en 1928, y posteriormente Vladimir K. Zworykin hizo lo mismo en 1929. Por el año de 1939, la British Broadcasting Corporation (BBC) transmitía televisión comercialmente.

En 1928, Harry Nyquist publicó un artículo clásico acerca de la teoría de la transmisión de señales en telegrafía. En particular, Nyquist creó criterios para la correcta recepción de señales telegráficas transmitidas por canales dispersos en la ausencia de ruido. Gran parte del trabajo inicial de Nyquist se aplicó después a la transmisión de datos digitales por canales dispersos.

En 1937, Alec Reeves inventó la *modulación por codificación de pulsos* (PCM) para la codificación digital de señales de voz. La técnica se creó durante la Segunda Guerra Mundial para permitir el encriptado de señales habladas; en realidad, el ejército estadounidense utilizó en el campo un sistema de 24 canales a escala máxima al final de la guerra. Sin embargo, la PCM tuvo que aguardar al descubrimiento del transistor y al subsecuente desarrollo de la integración a gran escala de circuitos para su explotación comercial.

En 1943, D.O. North ideó el *filtro acoplado* para la detección óptima de una señal conocida en ruido blanco aditivo. J.H. Van Vleck y D. Middleton obtuvieron en forma independiente un resultado similar en 1946; este último acuñó el término *filtro acoplado*.

En 1947, V.A. Kotel'nikov creó la representación geométrica de señales en una disertación doctoral presentada ante el Consejo Académico del Instituto de Energía Molotov en Moscú. John M. Wozencraft e Irwin M. Jacobs lograron después que este método fructificara completamente en un libro de texto memorable publicado en 1965.

En 1948, Claude Shannon estableció los fundamentos teóricos de las comunicaciones digitales en un artículo titulado "Una teoría matemática de la comunicación". Dicho artículo se recibió con una aclamación inmediata y entusiasta. Es posible que esta respuesta haya animado a Shannon a cambiar el título de su artículo por el de "La teoría matemática de la comunicación" cuando éste se reimprimió un año después en un libro en el que fue coautor Warren

Weaver. Vale la pena mencionar que antes de la publicación del artículo clásico de 1948 de Shannon, se pensaba que el aumento de la velocidad de transmisión de información sobre un canal incrementaría la probabilidad de error; la comunidad de la teoría de la comunicación se sorprendió cuando Shannon demostró que esto no era cierto, siempre que la velocidad de transmisión estuviera por debajo de la capacidad del canal. Al artículo de 1948 de Shannon siguieron algunos avances significativos en la teoría de la codificación, entre los cuales se incluyen los siguientes:

- Desarrollo de los primeros *códigos de corrección de errores* no triviales por parte de M.J.E. Golay en 1949 y de Richard W. Hamming en 1950.
- Desarrollo de los *códigos turbo* de C. Berrou, A. Glavieux y P. Thitimajshima en 1993; los códigos turbo proporcionan un desempeño en la codificación y en la decodificación de la corrección de errores casi óptimo en el sentido de Shannon.

En 1948, Walter H. Brattain, John Bardeen y William Shockley inventaron el *transistor* en los Laboratorios Bell. El primer circuito integrado (IC, por sus siglas en inglés) de silicio lo produjo Robert Noyce en 1958. Estas memorables innovaciones en los dispositivos de estado sólido y los circuitos integrados llevaron al desarrollo de los circuitos *integrados a muy grande escala de integración* (VLSI) y a los *microprocesadores* de un solo chip, y con ellos la naturaleza del procesamiento de señales y de la industria de las telecomunicaciones cambió para siempre.

La invención del transistor en 1948 estimuló la aplicación de la electrónica en las comunicaciones conmutables y en las digitales. La motivación era mejorar la confiabilidad, ampliar la capacidad y reducir los costos. La primera llamada a través de un sistema de programa almacenado ocurrió en marzo de 1958 en los Laboratorios Bell, y el primer servicio telefónico comercial con conmutación digital se inició en Morris, Illinois, en junio de 1960. La primera transmisión del *sistema de portadora T-1* se instaló en los Laboratorios Bell en 1962.

Durante el periodo de 1943 a 1946 se construyó la primera computadora digital electrónica, llamada ENIAC, en la Moore School of Electrical Engineering de la Universidad de Pensilvania, bajo la dirección técnica de J. Presper Eckert, Jr., y John W. Mauchly. Sin embargo, las contribuciones de John von Neumann fueron de las primeras y de las más fundamentales para la teoría, diseño y aplicación de las computadoras digitales, mismas que se remontan al primer borrador de un informe escrito en 1945. Las computadoras y las terminales empezaron a comunicarse entre sí a través de largas distancias a principios de la década de 1950. Los enlaces utilizados fueron inicialmente canales telefónicos de voz que operaban a bajas velocidades (300 a 1200 b/s). Diversos factores han contribuido a un aumento impresionante en las velocidades de transmisión de datos; entre ellos destaca la idea de la *ecualización adaptable*, que introdujo Robert Lucky en 1965, y técnicas de modulación eficiente, que presentó por primera vez G. Ungerboeck en 1982. Otra idea que se emplea ampliamente en las comunicaciones por computadora es la de *solicitud de repetición automática* (ARQ). El método ARQ lo ideó originalmente H.C.A. van Duuren durante la Segunda Guerra Mundial y se publicó en 1946. Se empleó para mejorar la radiotelefonía en la transmisión de télex a largas distancias.

Entre 1950 y 1970 se realizaron varios estudios acerca de *redes de computadoras*. No obstante, el más importante en términos del impacto en las comunicaciones por computadora fue el Advanced Research Project Agency Network (ARPANET), puesto en servicio por primera vez en 1971. El desarrollo de ARPANET lo patrocinó la Agencia de Proyectos de Investigación Avanzada del Departamento de Defensa de Estados Unidos. El trabajo pionero en la *conmutación en paquetes* se llevó a cabo en ARPANET. En 1985, esta última se renombró como *Internet*. El punto de viraje en la evolución de Internet ocurrió en 1990 cuando Tim Berners-Lee propuso para ella una interfaz de software hipermedia, a la que nombró *World Wide Web*.[12] Posteriormente, en el espacio de apenas casi dos años, la Web pasó de la nada a la popularidad mundial,

culminando en su comercialización en 1994. ¿Cómo explicar el explosivo crecimiento de Internet? Es posible responder a esta pregunta ofreciendo estas razones:[13]

- Antes de que la Web surgiera explosivamente, ya estaban listos los ingredientes para su creación. En particular, gracias a la VLSI, las computadoras personales ya se encontraban en gran parte de los hogares del planeta y cada vez mayor cantidad de ellas se equipaban con módems para la interconectividad con el mundo exterior.
- Por cerca de dos décadas, Internet había crecido uniformemente (aunque dentro de una comunidad confinada de usuarios), alcanzando un umbral crítico de correo electrónico y transferencia de archivos basado en el valor para el usuario.
- Se habían adoptado los estándares para la descripción y transferencia de documentos, el lenguaje de marcación de hipertexto (HTML) y el protocolo de transferencia de hipertexto (HTTP).

De ese modo, todo lo que era necesario para crear la Web ya estaba listo, con la excepción de dos ingredientes críticos: una interfaz de usuario simple y un concepto de servicio inteligente.

En 1955, John R. Pierce propuso el uso de satélites para las comunicaciones. Sin embargo, esta propuesta fue precedida por un artículo anterior de Arthur C. Clark que se publicó en 1945, donde se proponía también la idea de utilizar un satélite *orbital terrestre* como un punto repetidor para las comunicaciones entre dos estaciones terrestres. En 1957, la Unión Soviética lanzó el Sputnik I, que transmitió señales de telemetría durante 21 días. Al poco tiempo Estados Unidos lanzó, en 1958, el Explorer I, que transmitió señales de telemetría durante cerca de 5 meses. Se dio un gran paso experimental en la tecnología de los satélites de comunicaciones con el lanzamiento del Telstar I desde Cabo Cañaveral el 10 de julio de 1962. El satélite Telstar se había construido en los Laboratorios Bell, donde se había adquirido un conocimiento muy amplio a partir del trabajo pionero de Pierce. El satélite era capaz de transmitir programas de TV a través del Atlántico, lo cual sólo fue posible a través de usuarios de receptores máser y grandes antenas.

El empleo de medios ópticos (por ejemplo, señales de humo o de fuego) para la transmisión de información se remonta a los tiempos prehistóricos. Sin embargo, no hubo ningún acontecimiento notable en las comunicaciones ópticas hasta 1966, cuando K.C. Kao y G.A. Hockham de Standard Telephone Laboratories, en el Reino Unido, propusieron el uso de una fibra de vidrio revestida como una guía de ondas dieléctrica. El *láser* (acrónimo de amplificación de luz mediante emisión estimulada de radiación) se inventó y desarrolló en 1959 y 1960. Kao y Hockham señalaron que 1) la atenuación en una fibra óptica se debe a las impurezas del vidrio y que 2) la pérdida intrínseca, determinada por la dispersión de Rayleigh, es muy baja. En realidad, predijeron que era posible alcanzar una pérdida de 20 dB/km. Esta impresionante predicción, realizada en el tiempo en que las pérdidas de potencia en la fibra de vidrio eran aproximadamente de 1000 dB/km, se comprobó tiempo después. En nuestros días es posible conseguir pérdidas de transmisión tan bajas como 0.1 dB/km.

Los avances espectaculares en microelectrónica, computadoras digitales y sistemas de ondas luminosas que hemos atestiguado hasta ahora y que continuarán en el futuro, son causa de impresionantes cambios en el ambiente de las telecomunicaciones; muchos de estos cambios ya tienen carta de naturaleza y un mayor número de ellos surgirán en el tiempo que viene.

Notas y bibliografía

1. En lo que respecta a ensayos acerca de un recuento inicial de las comunicaciones y de otras disciplinas relacionadas (por ejemplo, electrónica, computadoras, radar, radioastronomía, satélites), consulte Overhage [112]; en particular vea el capítulo sobre "Comunicaciones" de L.V. Berkner, pp. 35-50.

2. El estándar de codificación de imágenes JPEG se explica en los artículos de Wallace [393]; vea también el artículo de T.A. Ramstad en el manual editado por Madisetti y Williams [99].
3. La transformada coseno discreta (DCT) y su inversa para un bloque de muestras de imagen de fuente de 8×8 se definen respectivamente por medio de

$$F(u,v) = \frac{1}{4}C(u)C(v)\left[\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2y+1)v\pi}{16}\right)\right]$$

$$f(x,y) = \frac{1}{4}\left[\sum_{u=0}^{7}\sum_{v=0}^{7} C(u)C(v)F(u,v)\cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2y+1)v\pi}{16}\right)\right]$$

donde

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{para } u = 0 \text{ y } v = 0 \\ 1 & \text{en otro caso} \end{cases}$$

Para un tratamiento completo de la DCT, vea Rao y Yip [125].

4. El estándar de codificación de video MPEG-1 se explica en el artículo de Gall [255]; vea también el artículo de A.M. Tekalp en el manual editado por Madisetti y Williams [99], el cual explica las versiones consecutivas del estándar de codificación de video MPEG.
5. El estándar de codificación de audio MPEG-1 se explica en los artículos de Brandenburg y Stoll [208] y Pan [336]; consulte también el artículo de P. Noll en el manual editado por Madisetti y Williams [99], que explica además las versiones consecutivas del estándar de codificación de audio MPEG. En particular, el uso extendido del estándar más actual, el de audio MPEG-3, está resultando en un nivel de piratería que quizá empequeñezca los problemas anteriores de las cintas de casetes "ilegales".
6. Para una explicación detallada de las redes de comunicación, vea Tanenbaum.
7. El modelo de referencia OSI fue creado por un subcomité de la Organización Internacional para la Estandarización (ISO) en 1977. En lo que respecta a una explicación de los principios implicados al llegar a las siete capas del modelo OSI y a una descripción de las mismas capas, vea Tanenbaum.
8. SONET fue propuesto originalmente por Telcordia Technologies Inc. (conocido entonces como Bellcore) y estandarizado por el Instituto Estadounidense de Estándares Nacionales (ANSI). Después, el Comité Consultivo Internacional para la Telefonía y la Telegrafía (CCITT) aprobó un estándar de la SONET y emitió un conjunto de recomendaciones simultáneas denominadas *jerarquía digital síncrona* (SDH). Las diferencias entre la SONET y la SDH son menores.
9. En lo que respecta a un análisis completo preciso de la propagación de las ondas luminosas en una fibra óptica, necesitamos tratar éstas como una guía de onda dieléctrica y utilizar las ecuaciones de Maxwell para llevar a cabo el análisis, el cual es de una naturaleza muy matemática. Para una explicación del análisis fácil de leer, vea el capítulo 3 de Green, Jr. [58].
10. En cuanto a un panorama poco técnico de los teoremas de Shannon acerca de la teoría de la información que se presentan en una forma muy accesible, vea el libro titulado *Silicon Dreams* de Lucky [96].
11. Para una descripción accesible de la historia de las comunicaciones, vea Lebow [84].
12. En lo que se refiere a una descripción histórica del desarrollo de Internet, consulte Leiner y otros [301].
13. En lo referente a un amplio ensayo acerca de los servicios de telecomunicaciones y a la forma en que la sociedad reacciona a su desarrollo, consulte Lucky. Este artículo señala la *ley de Metcalf*, de acuerdo con la cual aparentemente cualquier nuevo servicio de telecomunicaciones requiere un largo tiempo para que consiga la aceptación universal. Lucky cita a la World Wide Web como un contraejemplo asombroso para la ley de Metcalf y ofrece las razones de ello.

CAPÍTULO 1

PROCESOS ALEATORIOS

Este capítulo presenta un tratamiento introductorio de los procesos aleatorios estacionarios destacando la estadística de segundo orden. En particular, explica los siguientes aspectos:

- *La noción del proceso aleatorio.*
- *El requerimiento que debe satisfacerse para que un proceso aleatorio sea estacionario.*
- *La descripción parcial de un proceso aleatorio en términos de sus funciones: la media, la correlación y la covarianza.*
- *Las condiciones que tienen que satisfacerse para que un proceso aleatorio estacionario sea ergódico, una propiedad que nos permite sustituir los promedios del tiempo por promedios totales.*
- *¿Qué ocurre con un proceso aleatorio estacionario cuando se transmite a través de un filtro lineal invariante con el tiempo?*
- *La descripción en el dominio de la frecuencia de un proceso aleatorio en términos de la densidad espectral de potencia.*
- *Las características de un tipo importante de proceso aleatorio conocido como proceso gaussiano.*
- *Fuentes de ruido y su forma de banda angosta.*
- *Distribuciones de Rayleigh y Rician, las cuales representan dos distribuciones de probabilidad especiales que surgen en el estudio de los sistemas de comunicación.*

1.1 *Introducción*

La idea de un *modelo matemático* utilizado para describir un fenómeno físico se establece perfectamente en las ciencias físicas y en la ingeniería. En este contexto es posible distinguir dos clases de modelos matemáticos: determinístico y estocástico. Se afirma que un modelo es *determinístico* si no hay incertidumbre acerca de su comportamiento dependiente del tiempo en cualquier instante. Sin embargo, en muchos problemas del mundo real el empleo de un modelo determinístico resulta inapropiado porque el fenómeno físico de interés incluye demasiados factores desconocidos. A pesar de eso, es factible considerar un modelo descrito en términos probabilísticos en lo que señalamos como la *probabilidad* de un valor futuro que se encuentre entre dos límites especificados. En un caso de este tipo, se afirma que el modelo será *estocástico* o *aleatorio*. En el apéndice 1 se presenta un breve repaso de la teoría de probabilidad.

Considere, por ejemplo, un sistema de comunicación de radio. La señal recibida en dicho sistema suele constar de una componente de la *señal que lleva la información*, una componente de *interferencia* aleatoria y del *ruido del canal*. La componente de la señal que lleva la información tal vez represente, por ejemplo, una señal de voz que, por lo general consiste en ráfagas espaciadas al azar de energía de duración aleatoria. La componente de interferencia quizá represente ondas electromagnéticas espurias producidas por otros sistemas de comunicación que

operan en la vecindad del receptor de radio. Una fuente principal de ruido en el canal es el *ruido térmico*, que provoca el movimiento aleatorio de los electrones en conductores y dispositivos en el extremo frontal del receptor. De ese modo, descubrimos que la señal recibida es de naturaleza aleatoria. Aunque no es posible predecir desde el principio el valor exacto de la señal, sí es viable describirla en términos de parámetros estadísticos como la potencia promedio y la densidad espectral de potencia, como se explica en este capítulo.

1.2 *Definición matemática de un proceso aleatorio*

A la luz de estas notas introductorias, es claro que los procesos aleatorios tienen dos propiedades: 1) son funciones del tiempo y 2) son aleatorios porque antes de llevar a cabo un experimento, no es posible definir con exactitud las formas de onda que se observarán en el futuro.

Al describir un experimento aleatorio resulta conveniente pensar en términos de un espacio muestral. De modo específico, cada resultado del experimento se asocia con un *punto de la muestra*. La totalidad de los puntos de la muestra correspondientes al agregado de todos los resultados posibles del experimento recibe el nombre de *espacio muestral*. Cada punto de la muestra del espacio muestral es una función del tiempo. El espacio o totalidad muestral compuesta por funciones de tiempo se denomina *proceso aleatorio* o *estocástico*.[1] Como una parte integral de esta noción, suponemos la existencia de una distribución de probabilidad definida sobre una clase apropiada de conjuntos en el espacio muestral, por lo que hablamos con confianza de la probabilidad de diversos eventos.

Considere, entonces, un experimento aleatorio especificado por los resultados s de un *espacio muestral* S, por los eventos definidos en el espacio muestral S y por las probabilidades de

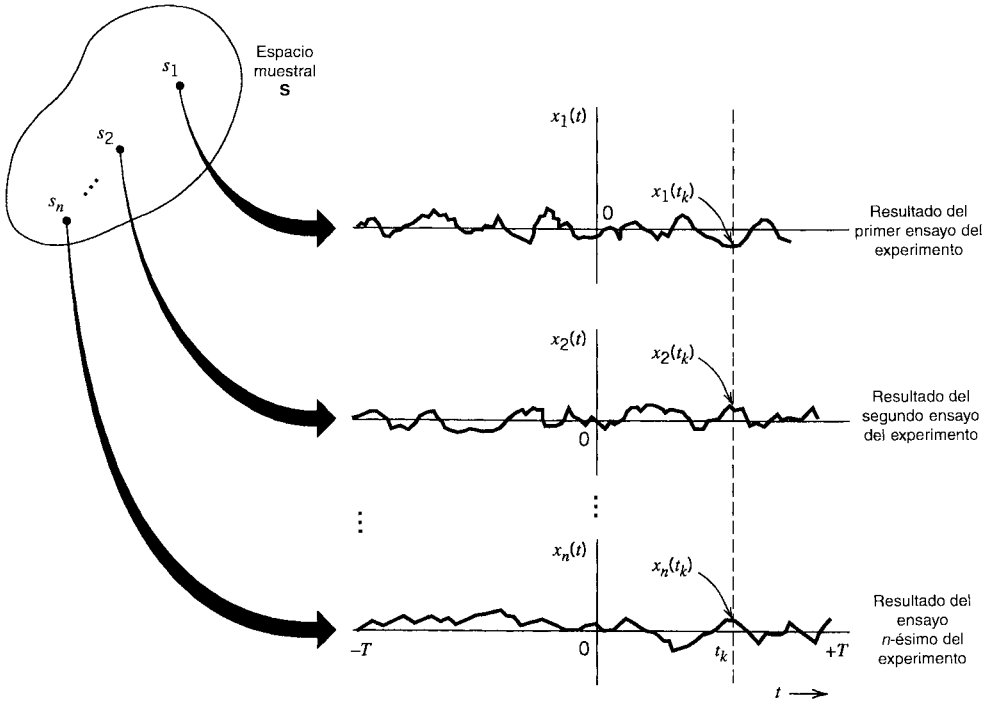


FIGURA 1.1 Un conjunto de funciones muestra.

estos eventos. Suponga que asignamos a cada punto s de la muestra una función del tiempo de acuerdo con la regla:

$$X(t, s), \qquad -T \leq t \leq T \tag{1.1}$$

donde $2T$ es el *intervalo de observación total.* En un punto fijo s_j de la muestra, la gráfica de la función $X(t, s_j)$ en función del tiempo t recibe el nombre de *realización* o *función de la muestra* del proceso aleatorio. Para simplificar la notación, denotamos esta función de la muestra como

$$x_j(t) = X(t, s_j) \tag{1.2}$$

La figura 1.1 ilustra un conjunto de funciones muestra $\{x_j(t) \mid j = 1, 2, \ldots, n\}$. A partir de esta figura, advertimos que para un tiempo fijo t_k dentro del intervalo de información, el conjunto de números

$$\{x_1(t_k), x_2(t_k), \ldots, x_n(t_k)\} = \{X(t_k, s_1)\ X(t_k, s_2). \ldots, X(t_k, s_n)\}$$

constituye una *variable aleatoria.* De esa manera, tenemos una totalidad (familia) indexada de variables aleatorias $\{X(t, s)\}$, que se denomina *proceso aleatorio.* Para simplificar la notación, una práctica común es suprimir la s y simplemente utilizar $X(t)$ para denotar un proceso aleatorio. En esas condiciones es posible definir en forma normal un proceso aleatorio $X(t)$ como *un conjunto de funciones de tiempo junto con una regla de probabilidad que asigna una probabilidad a cada evento significativo asociado con una observación de una de las funciones muestra del proceso aleatorio.* Asimismo, es posible distinguir entre una variable aleatoria y un proceso aleatorio de la manera siguiente:

- Para una variable aleatoria, el resultado de un experimento aleatorio se transforma en un número.
- En un proceso aleatorio, el resultado de un experimento aleatorio se transforma en una forma de onda que es una función del tiempo.

1.3 *Procesos estacionarios*

Al tratar con los procesos aleatorios que se encuentran en el mundo real, descubrimos muchas veces que la caracterización estadística del proceso es independiente del tiempo en el que se inicia la observación del proceso. Es decir, si un proceso de este tipo se divide en varios intervalos de tiempo, las distintas secciones del proceso exhiben en esencia las mismas propiedades estadísticas. Un proceso de estas características se dice que es *estacionario.* De otra manera, se indica que es *no estacionario.* En términos generales, un proceso estacionario surge de un fenómeno físico estable que ha evolucionado hasta un modo de comportamiento de estado estable, en tanto que un proceso no estacionario proviene de un fenómeno inestable.

Para ser más precisos, consideremos un proceso aleatorio $X(t)$ que se inicia en $t = -\infty$. Dejemos que $X(t_1), X(t_2), \ldots, X(t_k)$ denoten las variables aleatorias que se obtienen al observar los procesos aleatorios $X(t)$ en los tiempos $t_1, t_2, \ldots, t_k$, respectivamente. La función de distribución de unión de este conjunto de variables aleatorias es $F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k)$. Supóngase que a continuación desplazamos todas las observaciones en una cantidad fija τ, obteniendo de esa manera un nuevo conjunto de variables aleatorias $X(t_1 + \tau), X(t_2 + \tau), \ldots, X(t_k + \tau)$. La función de distribución de unión de este último conjunto de variables aleatorias es $F_{X(t_1 + \tau), \ldots, X(t_k + \tau)}(x_1, \ldots, x_k)$. Se dice que el proceso aleatorio $X(t)$ será *estacionario en el sentido estricto o estacionario estrictamente* si se cumple la siguiente condición:

$$F\hat{x}_{(t_1+\tau), \ldots X(t_k+\tau)}(x_1, \ldots, x_k) = F_{X(t_1), \ldots X(t_k)}(x_1, \ldots, x_k) \tag{1.3}$$

para todos los desplazamientos de tiempo τ, todas las k y todas las elecciones posibles del tiempo de observación $t_1, \ldots, t_k$. En otras palabras, un *proceso aleatorio* $X(t)$ iniciado en el tiempo $t = -\infty$, *es estrictamente estacionario si la distribución de unión de cualquier conjunto de variables aleatorias obtenidas al observar el proceso aleatorio* $X(t)$ *es invariante con respecto a la ubicación del origen* $t = 0$.

Advierta que las distribuciones de dimensión finita en la ecuación (1.3) dependen de la separación de tiempo relativa entre las variables aleatorias, pero no de su tiempo absoluto. Es decir, el proceso aleatorio tiene el mismo comportamiento probabilístico a través del tiempo.

De manera similar, podemos afirmar que dos procesos aleatorios $X(t)$ y $Y(t)$ son *en conjunto estrictamente estacionarios* si las distribuciones de unión de dimensión finita de los dos conjuntos de variables aleatorias $X(t_1), \ldots, X(t_k)$ y $Y(t'_1), \ldots, Yt'_j$ son invariables con respecto al origen $t = 0$ para todas las k y j y todas las elecciones de los tiempos de observación $t_1, \ldots, t_k$ y $t'_1, \ldots, t'_j$.

Al regresar a la ecuación (1.3), se distinguen dos situaciones de interés especial:

1. Para $k = 1$, tenemos

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x) \qquad \text{para todo } t \text{ y } \tau \tag{1.4}$$

Esto es, *la función de distribución de primer orden de un proceso aleatorio estacionario es independiente del tiempo.*

2. Para $k = 2$ y $\tau = -t_1$, tenemos

$$F_{X(t_1),X(t_2)}(x_1, x_2) = F_{X(0),X(t_2-t_1)}(x_1, x_2) \qquad \text{para todo } t_1 \text{ y } t_2 \tag{1.5}$$

Es decir, *la función de distribución de segundo orden de un proceso aleatorio estacionario sólo depende de la diferencia de tiempo entre los tiempos de observación* y no de los tiempos particulares a los cuales se observa el proceso aleatorio.

Estas dos propiedades tienen profundas implicaciones en la parametrización estadística de un proceso aleatorio estacionario; este aspecto se explica en la sección 1.4.

EJEMPLO 1.1

Considere la figura 1.2, que describe tres ventanas espaciales localizadas en los tiempos t_1, t_2, t_3. Deseamos evaluar la probabilidad de obtener una función muestra $x(t)$ de un proceso aleatorio $X(t)$ que pasa por este conjunto de ventanas; es decir, la probabilidad del evento de unión

$$A = \{a_i < X(t_i) \leq b_i\}, \quad i = 1, 2, 3$$

En términos de la función de distribución de unión, esta probabilidad es igual a

$$P(A) = F_{X(t_1),X(t_2),X(t_3)}(b_1, b_2, b_3) - F_{X(t_1),X(t_2),X(t_3)}(a_1, a_2, a_3)$$

Supongamos ahora que se sabe que el proceso aleatorio $X(t)$ es estrictamente estacionario. Una implicación de lo anterior es que la probabilidad de que el conjunto de funciones muestra de este proceso pase por las ventanas de la figura 1.3*a* es igual a la probabilidad de que el conjunto de funciones muestra pase por las ventanas correspondientes desplazadas en el tiempo de la figura 1.3*b*. Sin embargo, advierta que no es necesario que estos dos conjuntos estén compuestos por las mismas funciones muestra. ◀

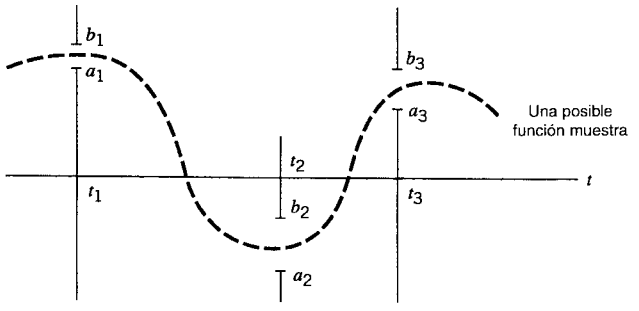


FIGURA 1.2 Ilustración de la probabilidad de un evento conjunto.

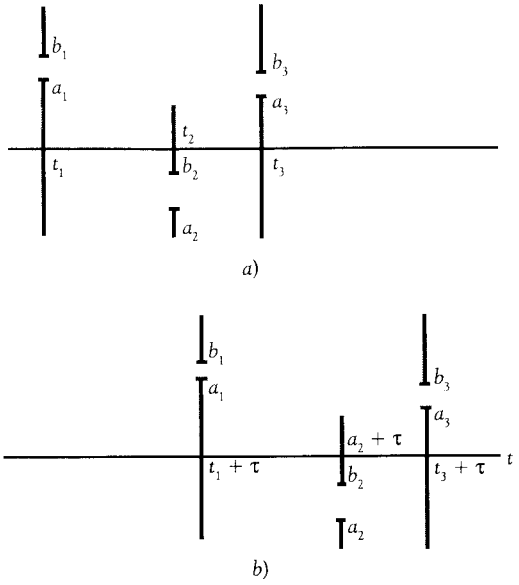


FIGURA 1.3 Ilustración del concepto de estacionariedad en el ejemplo 1.1.

1.4 *Funciones de la media, de correlación y de covarianza*

Considere un proceso aleatorio estrictamente estacionario $X(t)$. Definimos la *media* del proceso $X(t)$ como el valor esperado de la variable aleatoria obtenida al observar el proceso en algún tiempo t, como se indica mediante

$$\begin{aligned}\mu_X(t) &= E[X(t)] \\ &= \int_{-\infty}^{\infty} x f_{X(t)}(x)\, dx\end{aligned} \tag{1.6}$$

donde $f_{X(t)}(x)$ es la función de densidad de probabilidad de primer orden del proceso. A partir de la ecuación (1.4) deducimos que para un proceso aleatorio estrictamente estacionario, $f_{X(t)}(x)$ es independiente del tiempo t. En consecuencia, *la media de un proceso estrictamente estacionario es una constante*, como lo indica

$$\mu_X(t) = \mu_X \qquad \text{para todo } t \tag{1.7}$$

Definimos la *función de autocorrelación* del proceso $X(t)$ como el valor esperado del producto de dos variables aleatorias, $X(t_1)$ y $X(t_2)$, obtenido al observar el proceso $X(t)$ en los tiempos t_1 y t_2, respectivamente. De manera específica, escribimos

$$\begin{aligned}R_X(t_1, t_2) &= E[X(t_1)X(t_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1),X(t_2)}(x_1, x_2)\, dx_1 dx_2\end{aligned} \tag{1.8}$$

donde $f_{X(t_1),X(t_2)}(x_1, x_2)$ es la función de densidad de probabilidad de segundo orden del proceso. Según la ecuación (1.5), deducimos que para un proceso aleatorio estrictamente estacionario, $f_{X(t_1),X(t_2)}(x_1, x_2)$ sólo depende de la diferencia entre los tiempos de observación t_1 y t_2, lo cual

significa que la *función de autocorrelación de un proceso estrictamente estacionario sólo depende de la diferencia de tiempo* $t_2 - t_1$, como se indica mediante

$$R_X(t_1, t_2) = R_X(t_2 - t_1) \qquad \text{para todo } t_1 \text{ y } t_2 \tag{1.9}$$

De manera similar, la *función de autocovarianza* de un proceso estrictamente estacionario $X(t)$ se escribe como

$$C_X(t_1, t_2) = E[(X(t_1) - \mu_X)(X(t_2) - \mu_X)] \tag{1.10}$$
$$= R_X(t_2 - t_1) - \mu_X^2$$

La ecuación (1.10) muestra que, al igual que la función de autocorrelación, la función de autocovarianza de un proceso estrictamente estacionario $X(t)$ sólo depende de la diferencia del tiempo $t_2 - t_1$. Esta ecuación indica también que si conocemos la media y la función de autocorrelación del proceso, es posible que determinemos en forma única la función de autocovarianza. La media y la función de autocorrelación son, por tanto, suficientes para describir los primeros dos momentos del proceso.

Sin embargo, deben hacerse notar con todo cuidado dos puntos importantes:

1. La media y la función de autocorrelación proporcionan sólo una *descripción parcial* de la distribución de un proceso aleatorio $X(t)$.
2. Las condiciones de las ecuaciones (1.7) y (1.9), que significan la media y la función de autocorrelación respectivamente, *no* son suficientes para garantizar que el proceso aleatorio $X(t)$ sea estrictamente estacionario.

En consecuencia, las consideraciones prácticas señalan que muchas veces nos limitamos simplemente a una descripción parcial del proceso dado por la media y la función de autocorrelación. La clase de procesos aleatorios que satisfacen las ecuaciones (1.7) y (1.9) reciben diversos nombres: *procesos estacionarios de segundo orden, estacionarios de sentido amplio* o *débilmente estacionarios*. De aquí en adelante, nos referiremos a ellos simplemente como *procesos estacionarios*.[2]

Un proceso estacionario no es por obligación estrictamente estacionario, ya que las ecuaciones (1.7) y (1.9) evidentemente no implican la invarianza de la distribución de unión (k dimensional) de la ecuación (1.3) con respecto al desplazamiento de tiempo τ para toda k. Por otro lado, un proceso estrictamente estacionario no siempre satisface las ecuaciones (1.7) y (1.9), pues los momentos de primero y segundo orden quizá no existan. Sin embargo, es claro que la clase de procesos estrictamente estacionarios con momentos finitos de segundo orden forma una subclase de la clase de todos los procesos estacionarios.

■ PROPIEDADES DE LA FUNCIÓN DE AUTOCORRELACIÓN

Por conveniencia de la notación, redefinimos la función de autocorrelación de un proceso estacionario $X(t)$ como

$$R_X(\tau) = E[X(t + \tau)X(t)] \qquad \text{para todo } t \tag{1.11}$$

Esta función de autocorrelación tiene varias propiedades importantes:

1. El *valor cuadrático medio* del proceso se puede obtener a partir de $R_X(\tau)$ poniendo simplemente $\tau = 0$ en la ecuación (1.11), como se indica por medio de

$$R_X(0) = E[X^2(t)] \tag{1.12}$$

2. La función de autocorrelación $R_X(\tau)$ es una función par de τ, esto es,

$$R_X(\tau) = R_X(-\tau) \tag{1.13}$$

Esta propiedad se desprende directamente de la ecuación de definición (1.11). En consecuencia, es posible que definamos también la función de autocorrelación $R_X(\tau)$ como

$$R_X(\tau) = E[X(t)X(t - \tau)]$$

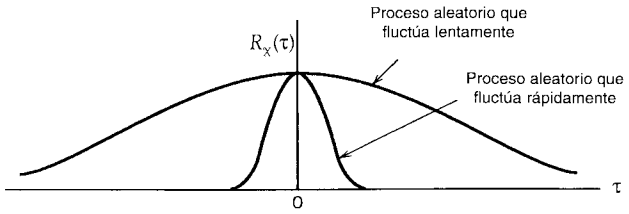


FIGURA 1.4 Ilustración de las funciones de autocorrelación de procesos aleatorios que fluctúan lenta y rápidamente.

3. La función de autocorrelación $R_X(\tau)$ tiene su magnitud máxima en $\tau = 0$, es decir,

$$|R_X(\tau)| \leq R_X(0) \tag{1.14}$$

Para probar esta propiedad, considere esta cantidad no negativa

$$E[(X(t+\tau) \pm X(t))^2] \geq 0$$

Al desarrollar los términos y tomar sus valores esperados individuales, encontramos fácilmente que

$$E[X^2(t+\tau)] \pm 2E[X(t+\tau)X(t)] + E[X^2(t)] \geq 0$$

que, a la luz de las ecuaciones (1.11) y (1.12) se reduce a

$$2R_X(0) \pm 2R_X(\tau) \geq 0$$

De manera equivalente, es posible escribir

$$-R_X(0) \leq R_X(\tau) \leq R_X(0)$$

a partir de lo cual se deduce directamente la ecuación (1.14).

El significado físico de la función de autocorrelación $R_X(\tau)$ es que ésta proporciona un medio para describir la interdependencia de dos variables aleatorias obtenidas al observar un proceso aleatorio $X(t)$ en tiempo separado τ segundos. Por tanto, resulta claro que cuanto más rápido cambie con el tiempo el proceso aleatorio $X(t)$, tanto más rápido la función de correlación $R_X(\tau)$ disminuye a partir de su máximo $R_X(0)$ cuando τ crece, como se ilustra en la figura 1.4. Esta definición tal vez se caracterice por medio de un *tiempo de correlación inversa* τ_0, tal que para $\tau > \tau_0$, la magnitud de la función de autocorrelación $R_X(\tau)$ permanece por debajo de algún valor preestablecido. De tal modo, podemos definir el tiempo de correlación inversa τ_0 del proceso estacionario $X(t)$ de media cero como el tiempo que se requiere para que la magnitud de la función de autocorrelación $R_X(\tau)$ disminuya hasta 1%, digamos, de su valor máximo $R_X(0)$.

EJEMPLO 1.2 Onda senoidal con fase aleatoria

Considere una señal senoidal con fase aleatoria, definida por

$$X(t) = A\cos(2\pi f_c t + \Theta) \tag{1.15}$$

donde A y f_c son constantes y Θ es una variable aleatoria que está *distribuida uniformemente* sobre el intervalo $[-\pi, \pi]$, esto es,

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & -\pi \leq \theta \leq \pi \\ 0, & \text{en otra parte} \end{cases} \tag{1.16}$$

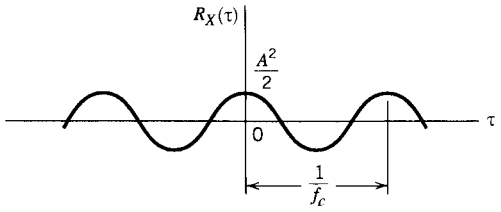


FIGURA 1.5 Función de autocorrelación de una onda seno con fase aleatoria.

Lo cual quiere decir que es igualmente probable que la variable Θ tenga cualquier valor θ en el intervalo $[-\pi, \pi]$. Cada valor de Θ corresponde a una muestra en el espacio muestral en el proceso aleatorio $X(t)$.

El proceso $X(t)$ definido por las ecuaciones (1.15) y (1.16) quizá represente una portadora generada localmente en el receptor de un sistema de comunicaciones, que se usa en la demodulación en la señal recibida. En particular, la variable aleatoria Θ denota la diferencia de fase entre esta portadora generada localmente y la onda portadora senoidal que se utilizó para modular la señal del mensaje en el transmisor.

La función de autocorrelación de $X(t)$ es

$$\begin{aligned} R_X(\tau) &= E[X(t+\tau)X(t)] \\ &= E[A^2 \cos(2\pi f_c t + 2\pi f_c \tau + \Theta)\cos(2\pi f_c t + \Theta)] \\ &= \frac{A^2}{2}E[\cos(4\pi f_c t + 2\pi f_c \tau + 2\Theta)] + \frac{A^2}{2}E[\cos(2\pi f_c \tau)] \\ &= \frac{A^2}{2}\int_{-\pi}^{\pi}\frac{1}{2\pi}\cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)\,d\theta + \frac{A^2}{2}\cos(2\pi f_c \tau) \end{aligned}$$

La integral del primer término es igual a cero, por lo que obtenemos

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau) \tag{1.17}$$

que se grafica en la figura 1.5. En consecuencia, vemos que la función de autocorrelación de una onda senoidal con fase aleatoria es otra senoide a la misma frecuencia en el "dominio de τ" en vez del dominio de tiempo original. ◀

▶ EJEMPLO 1.3 Onda binaria aleatoria

La figura 1.6 representa la función de muestra $x(t)$ de un proceso $X(t)$ consistente en una secuencia aleatoria de *símbolo binario* 1 y 0. Se hacen las siguientes suposiciones:

1. Los símbolos 1 y 0 se representan mediante pulsos de amplitud $+A$ y $-A$ volts, respectivamente, y de T segundos de duración.
2. Los pulsos no están sincronizados, por lo que el tiempo de inicio t_d del primer pulso completo para el tiempo positivo es igualmente probable que se encuentre entre 0 y T segundos. Esto es, t_d es el valor de muestra de una variable aleatoria distribuida uniformemente T_d, con su función de densidad de probabilidad definida por

$$f_{T_d}(t_d) = \begin{cases} \frac{1}{T}, & 0 \le t_d \le T \\ 0, & \text{en otra parte} \end{cases}$$

3. Durante cualquier intervalo de tiempo $(n-1)T < t - t_d < nT$, donde n es un entero, la presencia de un 1 o un 0 se determina lanzando una moneda sin alteraciones; específicamente, si el resultado es cara, tenemos un 1, y si lo que resulta es cruz, tenemos un 0. De ese modo, estos dos

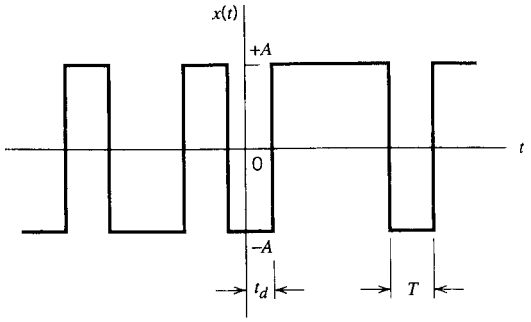


FIGURA 1.6 Función muestra de una onda binaria aleatoria.

símbolos son igualmente probables y la presencia de un 1 o un 0 en cualquier intervalo es independiente de todos los demás intervalos.

Puesto que los niveles de amplitud $-A$ y $+A$ ocurren con igual probabilidad, se concluye de inmediato que $E[X(t)] = 0$ para todo t, y la media del proceso resulta consecuentemente 0.

Para determinar la función de autocorrelación $R_X(t_k, t_i)$, tenemos que evaluar $E[X(t_k)X(t_i)]$, donde $X(t_k)$ y $X(t_i)$ son variables aleatorias obtenidas al observar el proceso aleatorio $X(t)$ en los tiempos t_k y t_i, respectivamente.

Considere primero el caso en el que $|t_k - t_i| > T$. Bajo esta condición las variables aleatorias $X(t_k)$ y $X(t_i)$ ocurren en diferentes intervalos de pulso y son, por lo tanto, independientes. De tal modo tenemos

$$E[X(t_k)X(t_i)] = E[X(t_k)]E[X(t_i)] = 0, \quad |t_k - t_i| > T$$

Considere a continuación el caso en el que $|t_k - t_i| < T$, con $t_k = 0$ y $t_i < t_k$. En una situación de este tipo observamos en la figura 1.6 que las variables aleatorias $X(t_k)$ y $X(t_i)$ ocurren en el mismo intervalo de pulso si y sólo si el retraso t_d satisface la condición $t_d < T - |t_k - t_i|$. Así, obtenemos la *esperanza condicional:*

$$E[X(t_k)X(t_i)|t_d] = \begin{cases} A^2, & t_d < T - |t_k - t_i| \\ 0, & \text{en otra parte} \end{cases}$$

Al promediar este resultado sobre todos los valores posibles de t_d, obtenemos

$$\begin{aligned} E[X(t_k)X(t_i)] &= \int_0^{T-|t_k-t_i|} A^2 f_{T_d}(t_d)\, dt_d \\ &= \int_0^{T-|t_k-t_i|} \frac{A^2}{T} dt_d \\ &= A^2\left(1 - \frac{|t_k - t_i|}{T}\right), \quad |t_k - t_i| < T \end{aligned}$$

Mediante un razonamiento similar de otro valor de t_k, concluimos que la función de autocorrelación de una onda binaria aleatoria, representada por la función muestra que se presenta en la figura 1.6, es sólo una función en la diferencia de tiempo $\tau = t_k - t_i$, como se indica por medio de

$$R_X(\tau) = \begin{cases} A^2\left(1 - \dfrac{|\tau|}{T}\right), & |\tau| < T \\ 0, & |\tau| \geq T \end{cases} \tag{1.18}$$

Este resultado se grafica en la figura 1.7. ◀

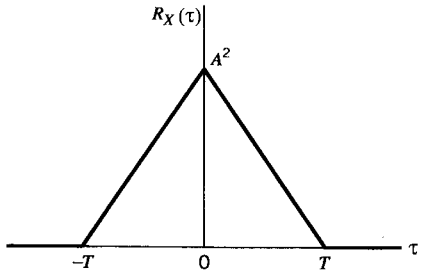


FIGURA 1.7 Función de autocorrelación de una onda binaria aleatoria.

■ FUNCIONES DE CORRELACIÓN CRUZADA

Considere a continuación el caso más general de dos procesos aleatorios $X(t)$ y $Y(t)$ con funciones de autocorrelación $R_X(t, u)$ y $R_Y(t, u)$, respectivamente. Las dos *funciones de correlación cruzada* de $X(t)$ y $Y(t)$ se definen por medio de

$$R_{XY}(t,u) = E[X(t)Y(u)] \tag{1.19}$$

y

$$R_{YX}(t,u) = E[Y(t)X(u)] \tag{1.20}$$

donde t y u denotan dos valores de tiempo en los cuales se observan los procesos. En este caso, las propiedades de correlación de los dos procesos aleatorios $X(t)$ y $Y(t)$ quizá se exhiban de manera conveniente en forma de matriz del siguiente modo:

$$\mathbf{R}(t,u) = \begin{bmatrix} R_X(t,u) & R_{XY}(t,u) \\ R_{YX}(t,u) & R_Y(t,u) \end{bmatrix}$$

la cual recibe el nombre de *matriz de correlación* del proceso aleatorio $X(t)$ y $Y(t)$. Si cada uno de los procesos aleatorios $X(t)$ y $Y(t)$ es estacionario y, además, son conjuntamente estacionarios, entonces es posible escribir la matriz de correlación como

$$\mathbf{R}(\tau) = \begin{bmatrix} R_X(\tau) & R_{XY}(\tau) \\ R_{YX}(\tau) & R_Y(\tau) \end{bmatrix} \tag{1.21}$$

donde $\tau = t - u$.

La función de correlación cruzada no es en general una función por par de τ, como sí ocurrió para la función de autocorrelación, y tiene un máximo en el origen. Sin embargo, obedece a cierta relación de simetría como la siguiente (véase el problema 1.9):

$$R_{XY}(\tau) = R_{YX}(-\tau) \tag{1.22}$$

▶ EJEMPLO 1.4 Procesos modulados en cuadratura

Considere un par de *procesos modulados en cuadratura* $X_1(t)$ y $X_2(t)$ que se relacionan con un proceso estacionario $X(t)$ como sigue:

$$X_1(t) = X(t)\cos(2\pi f_c t + \Theta)$$
$$X_2(t) = X(t)\,\text{sen}(2\pi f_c t + \Theta)$$

donde f_c es una frecuencia de portadora y la variable aleatoria Θ está distribuida uniformemente sobre el intervalo $[0, 2\pi]$. Además, Θ es independiente de $X(t)$. Una función de correlación cruzada de $X_1(t)$ y $X_2(t)$ está dada por

$$\begin{aligned} R_{12}(\tau) &= E[X_1(t)X_2(t-\tau)] \\ &= E[X(t)X(t-\tau)\cos(2\pi f_c t + \Theta)\,\text{sen}\,(2\pi f_c t - 2\pi f_c \tau + \Theta)] \\ &= E[X(t)X(t-\tau)]E[\cos(2\pi f_c t + \Theta)\,\text{sen}(2\pi f_c t - 2\pi f_c \tau + \Theta)] \\ &= \tfrac{1}{2}R_X(\tau)E[\text{sen}\,(4\pi f_c t - 2\pi f_c \tau + 2\Theta) - \text{sen}(2\pi f_c \tau)] \\ &= -\tfrac{1}{2}R_X(\tau)\ \text{sen}(2\pi f_c \tau) \end{aligned} \tag{1.23}$$

donde, en la última línea, hemos hecho uso de la distribución uniforme de la variable aleatoria Θ que representa a la fase. Advierta que en $\tau = 0$, el factor $\text{sen}(2\pi f_c \tau)$ es 0 y, en consecuencia,

$$\begin{aligned} R_{12}(0) &= E[X_1(t)X_2(t)] \\ &= 0 \end{aligned}$$

Esto demuestra que las variables aleatorias obtenidas mediante la observación simultánea de procesos de modelación en cuadratura $X_1(t)$ y $X_2(t)$ en algún valor fijo de tiempo t son ortogonales entre sí. ◀

1.5 *Procesos ergódicos*

Las *esperanzas* o *promedios totales* de un proceso aleatorio $X(t)$ son promedios "entre los extremos del proceso". Por ejemplo, la media de un proceso aleatorio $X(t)$ en algún tiempo fijo t_k es la esperanza de la variable aleatoria $X(t_k)$ que describe *todos los valores posibles* de las funciones muestra del proceso observado en el tiempo $t = t_k$. De modo natural, es posible definir también *promedios de muestra a largo plazo* o *promedios de tiempo* que son promedio "a lo largo del proceso". Por lo tanto, estamos interesados en relacionar promedios totales con promedios de tiempo para que estos últimos representen un medio práctico que esté disponible para la *estimación* de los promedios totales de un proceso aleatorio. La pregunta clave, desde luego, es: ¿Cuándo es posible sustituir los promedios de tiempo por los promedios totales? Para explorar este aspecto, considere la función muestra $x(t)$ de un proceso estacionario $X(t)$, con el intervalo de observación definido como $-T \le t \le T$. El *valor* de cd de $x(t)$ se define mediante el promedio de tiempo

$$\mu_x(T) = \frac{1}{2T}\int_{-T}^{T} x(t)\,dt \tag{1.24}$$

Claramente, el promedio de tiempo $\mu_x(T)$ es una variable aleatoria, ya que su valor depende del intervalo de observación y de qué función muestra particular del proceso aleatorio $X(t)$ se toma para utilizarse en la ecuación (1.24). Puesto que se supone que el proceso $X(t)$ es estacionario, la media del promedio de tiempo $\mu_x(T)$ está dada por (después de intercambiar la operación de esperanza e integración):

$$\begin{aligned} E[\mu_x(T)] &= \frac{1}{2T}\int_{-T}^{T} E[x(t)]\,dt \\ &= \frac{1}{2T}\int_{-T}^{T} \mu_X\,dt \\ &= \mu_X \end{aligned} \tag{1.25}$$

donde μ_X es la media del proceso $X(t)$. En consecuencia, el promedio de tiempo $\mu_x(T)$ representa una *estimación no polarizada* de la media promediada en el conjunto μ_X. Afirmamos que el proceso $X(t)$ es *ergódico en la media* si se satisfacen dos condiciones:

- El promedio de tiempo $\mu_x(T)$ se acerca al promedio total μ_X en el límite cuando el intervalo de observación T tiende a infinito; esto es,

$$\lim_{T\to\infty} \mu_x(T) = \mu_X$$

- La varianza de $\mu_x(T)$, tratada como una variable aleatoria, tiende a 0 en el límite cuando el intervalo de observación T se acerca a infinito; esto es,

$$\lim_{T\to\infty} \text{var}[\mu_x(T)] = 0$$

El otro promedio en el tiempo de interés particular es la función de autocorrelación $R_x(\tau, T)$ definido en términos de la función muestra $x(t)$ observada sobre el intervalo $-T \leq t \leq T$. Siguiendo la ecuación (1.24) es posible que definamos formalmente la *función de autocorrelación promediada en el tiempo* de una función muestra $x(t)$ del modo siguiente:

$$R_x(\tau, T) = \frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)\,dt \tag{1.26}$$

Este segundo promedio en el tiempo también debe verse como una variable aleatoria con una media y una valencia propias. De manera similar a la ergodicidad de la media, afirmamos que el proceso $x(t)$ es *ergódico en la función de autocorrelación* si se satisfacen las dos condiciones límite siguientes:

$$\lim_{T\to\infty} R_x(\tau, T) = R_X(\tau)$$

$$\lim_{T\to\infty} \text{var}[R_x(\tau, T)] = 0$$

En realidad, podríamos proseguir de modo similar para definir la ergodicidad en el sentido más general al considerar la estadística de orden superior del proceso $X(t)$. Sin embargo, en la práctica la ergodicidad de la media y la ergodicidad en la función de autocorrelación, como se describen aquí, se consideran a menudo (aunque no siempre) adecuadas. Advierta también que el uso de las ecuaciones (1.24) y (1.26) para calcular los promedios de tiempo $\mu_x(T)$ y $R_x(t, T)$ requieren que el proceso $X(t)$ sea estacionario. En otras palabras, para que un proceso aleatorio sea ergódico, tiene que ser estacionario; sin embargo, lo inverso no es necesariamente cierto.

1.6 *Transmisión de un proceso aleatorio a través de un filtro lineal invariante con el tiempo*

Suponga que un proceso aleatorio $X(t)$ se aplica como entrada a un *filtro lineal invariante con el tiempo de respuesta al impulso* $h(t)$, produciendo un nuevo proceso aleatorio $Y(t)$ en la salida del filtro, como en la figura 1.8. En general, resulta difícil describir la distribución de probabilidad del proceso aleatorio de salida $Y(t)$, aun cuando la distribución de probabilidad del proceso aleatorio de entrada $X(t)$ esté especificada por completo en $-\infty < t < \infty$.

En esta sección determinamos la forma en el dominio del tiempo de las relaciones de entrada-salida del filtro para definir la media y las funciones de autocorrelación del proceso aleatorio de salida $Y(t)$ en términos de las correspondientes a la entrada $X(t)$, suponiendo que $X(t)$ es un proceso estacionario.

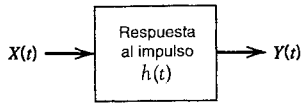


FIGURA 1.8 Transmisión de un proceso aleatorio a través de un filtro lineal invariante con el tiempo.

La transmisión de un proceso a través de un filtro lineal invariante con el tiempo está gobernada por la *integral de convolución*; para un repaso de esta operación, consulte el apéndice 2. En lo que respecta al problema presente, es posible que expresemos el proceso aleatorio de salida $Y(t)$ en términos del proceso aleatorio de entrada $X(t)$ como

$$Y(t) = \int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1$$

donde τ_1 es la variable de integración. Por consiguiente, la media de $Y(t)$ es

$$\begin{aligned}\mu_Y(t) &= E[Y(t)] \\ &= E\left[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1\right]\end{aligned} \tag{1.27}$$

Siempre que la esperanza $E[X(t)]$ sea finita para todo t y el sistema sea estable, existe la posibilidad de que intercambiemos el orden de la esperanza y la integración en la ecuación (1.27) y escribir de esa manera

$$\begin{aligned}\mu_Y(t) &= \int_{-\infty}^{\infty} h(\tau_1)E[X(t-\tau_1)]\,d\tau_1 \\ &= \int_{-\infty}^{\infty} h(\tau_1)\mu_X(t-\tau_1)\,d\tau_1\end{aligned} \tag{1.28}$$

Cuando el proceso aleatorio de entrada $X(t)$ es estacionario, la media $\mu_X(t)$ es una constante μ_X, por lo que podemos simplificar la ecuación (1.28) del modo siguiente:

$$\begin{aligned}\mu_Y &= \mu_X \int_{-\infty}^{\infty} h(\tau_1)\,d\tau_1 \\ &= \mu_X H(0)\end{aligned} \tag{1.29}$$

donde $H(0)$ es la respuesta en frecuencia 0 (cd) del sistema. La ecuación (1.29) establece que la media de un proceso aleatorio $Y(t)$ producida a la salida de un sistema lineal invariante con el tiempo en respuesta a $X(t)$ que actúa como el proceso de entrada es igual a la media de $X(t)$ multiplicada por la respuesta de cd del sistema, lo cual se satisface de manera intuitiva.

Considere la función de autocorrelación del proceso aleatorio de salida $Y(t)$. Por definición, tenemos

$$R_Y(t,u) = E[Y(t)Y(u)]$$

donde t y u denotan dos valores del tiempo en el cual se observa el proceso de salida. Por tanto, resulta viable utilizar la integral de convolución para escribir

$$R_Y(t,u) = E\left[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)X(u-\tau_2)\,d\tau_2\right] \tag{1.30}$$

En este caso también, siempre que el valor cuadrático medio $E[X^2(t)]$ sea finito para todo t y el sistema sea estable, podemos intercambiar el orden de la esperanza y la integración con respecto a τ_1 y τ_2 en la ecuación (1.30), obteniendo

$$\begin{aligned} R_Y(t,u) &= \int_{-\infty}^{\infty} d\tau_1 h(\tau_1) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) E[X(t-\tau_1)X(u-\tau_2)] \\ &= \int_{-\infty}^{\infty} d\tau_1 h(\tau_1) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) R_X(t-\tau_1, u-\tau_2) \end{aligned} \tag{1.31}$$

Cuando la entrada $X(t)$ es un proceso estacionario, la función de autocorrelación de $X(t)$ es únicamente una función de la diferencia entre los tiempos de observación $t - \tau_1$ y $u - \tau_2$. Por consiguiente, al poner $\tau = t - u$ en la ecuación (1.31), es posible escribir

$$R_Y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1) h(\tau_2) R_X(\tau - \tau_1 + \tau_2)\, d\tau_1 d\tau_2 \tag{1.32}$$

Al combinar este resultado con el que involucra a la media μ_Y, observemos que si la entrada para un filtro lineal y estable invariante con el tiempo es un proceso estacionario, entonces la salida del filtro también es un proceso estacionario.

Puesto que $R_Y(0) = E[Y^2(t)]$, se concluye que el valor medio cuadrático del proceso aleatorio de salida $Y(t)$ se obtiene al sustituir $\tau = 0$ en la ecuación (1.32). De tal modo obtenemos el resultado

$$E[Y^2(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1) h(\tau_2) R_X(\tau_2 - \tau_1)\, d\tau_1\, d\tau_2 \tag{1.33}$$

que es una constante.

1.7 *Densidad espectral de potencia*

Hasta ahora hemos considerado la caracterización de procesos estacionarios en sistemas lineales en el dominio del tiempo. A continuación nos enfocaremos a la caracterización de procesos aleatorios en sistemas lineales utilizando ideas en el dominio de la frecuencia. En particular, deseamos reducir el equivalente en el dominio de la frecuencia del resultado de la ecuación (1.33) definiendo el valor medio cuadrático de la salida del filtro.

Por definición, la respuesta al impulso de un filtro invariante con el tiempo es igual a la transformada inversa de Fourier de la respuesta en frecuencia del sistema; un repaso de la transformada de Fourier se presenta en el apéndice 2. Al utilizar $H(f)$ para denotar la *respuesta en frecuencia* del sistema, es posible escribir

$$h(\tau_1) = \int_{-\infty}^{\infty} H(f) \exp(j2\pi f \tau_1)\, df \tag{1.34}$$

Mediante la sustitución de esta expresión para $h(\tau_1)$ en la ecuación (1.33), obtenemos

$$\begin{aligned} E[Y^2(t)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} H(f) \exp(j2\pi f \tau_1) df \right] h(\tau_2) R_X(\tau_2 - \tau_1)\, d\tau_1\, d\tau_2 \\ &= \int_{-\infty}^{\infty} df\, H(f) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) \int_{-\infty}^{\infty} R_X(\tau_2 - \tau_1) \exp(j2\pi f \tau_1)\, d\tau_1 \end{aligned} \tag{1.35}$$

En la última integral del segundo miembro de la ecuación (1.35) definimos una nueva variable

$$\tau = \tau_2 - \tau_1$$

La ecuación (1.35) puede reescribirse después en la forma

$$E[Y^2(t)] = \int_{-\infty}^{\infty} df H(f) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) \exp(j2\pi f\tau_2) \int_{-\infty}^{\infty} R_X(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.36}$$

Sin embargo, la integral intermedia del segundo miembro de la ecuación (1.36) es simplemente $H^*(f)$, la conjugada compleja de la respuesta en frecuencia del filtro, y por ello esta ecuación puede simplificarse como

$$E[Y^2(t)] = \int_{-\infty}^{\infty} df |H(f)|^2 \int_{-\infty}^{\infty} R_X(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.37}$$

donde $|H(f)|$ es la *respuesta en magnitud* del filtro. La ecuación se simplifica adicionalmente al reconocer que la última integral es simplemente la transformada de Fourier de la función de autocorrelación $R_X(\tau)$ del proceso aleatorio de entrada $X(t)$. Esto nos lleva a introducir la definición de un nuevo parámetro

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.38}$$

La función $S_X(f)$ recibe el nombre de *densidad espectral de potencia* o *espectro de potencia,* del proceso estacionario $X(t)$. De modo que al sustituir la ecuación (1.38) en la (1.37), obtenemos la relación deseada:

$$E[Y^2(t)] = \int_{-\infty}^{\infty} |H(f)|^2 S_X(f)\, df \tag{1.39}$$

La ecuación (1.39) establece que *el valor cuadrático medio de la salida de un filtro lineal invariante con el tiempo en respuesta a un proceso estacionario es igual a la integral sobre todas las frecuencias de la densidad espectral de potencia del proceso de entrada multiplicado por la respuesta en magnitud al cuadrado del filtro.* Éste es el equivalente deseado en el dominio de la frecuencia para la relación en el dominio del tiempo en la ecuación (1.33).

Para indagar el significado físico de la densidad espectral de potencia, suponga que el proceso aleatorio $X(t)$ se hace pasar a través de un filtro de banda angosta ideal con respuesta en magnitud centrada en torno a la frecuencia f_c, como se indica en la figura 1.9; esto es,

$$|H(f)| = \begin{cases} 1, & |f \pm f_c| < \frac{1}{2}\Delta f \\ 0, & |f \pm f_c| > \frac{1}{2}\Delta f \end{cases} \tag{1.40}$$

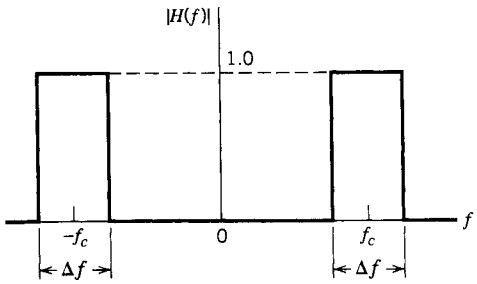


FIGURA 1.9 Respuesta en magnitud de un filtro de banda angosta ideal.

donde Δf es el ancho de banda del filtro. Entonces, de acuerdo con la ecuación (1.39), encontramos que si el ancho de banda del filtro Δf es suficientemente pequeño comparado con las frecuencias de la banda media f_c y $S_X(f)$ es una función continua, el valor cuadrático medio de la salida del filtro es aproximadamente

$$E[Y^2(t)] \simeq (2\Delta f) S_X(f_c) \tag{1.41}$$

Sin embargo, el filtro deja pasar sólo aquellas componentes de frecuencia del proceso aleatorio de entrada $X(t)$ que están dentro de una banda de frecuencias estrecha de ancho Δf de entrada en torno a la frecuencia $\pm f_c$. De esa manera, $S_X(f_c)$ representa la densidad de frecuencia de la potencia promedio en el proceso aleatorio $X(t)$, evaluada a la frecuencia $f = f_c$. Las dimensiones de la densidad espectral de frecuencia están, por tanto, en watts por Hertz (W/Hz).

■ PROPIEDADES DE LA DENSIDAD ESPECTRAL DE POTENCIA

La densidad espectral de potencia $S_X(f)$ y la función de autocorrelación $R_X(\tau)$ de un proceso estacionario $X(t)$ forman un par de transformadas de Fourier τ y f como las variables de interés, según indica el par de relaciones

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.42}$$

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) \exp(j2\pi f\tau)\, df \tag{1.43}$$

Las ecuaciones (1.42) y (1.43) son relaciones básicas en la teoría del análisis espectral de procesos aleatorios, y en conjunto constituyen lo que suele denominarse *relaciones de Einstein-Wiener-Khintchine.*[3]

Dichas relaciones muestran que si se conoce la función de autocorrelación con la densidad espectral de potencia de un proceso aleatorio, la otra puede determinarse con exactitud. Sin embargo, estas funciones exhiben diferentes aspectos de la información de correlación en torno al proceso. Se acepta comúnmente que para procesos prácticos, la densidad espectral de potencia es el "parámetro" de mayor utilidad.

A continuación deseamos utilizar este par de relaciones para obtener algunas propiedades generales de la densidad espectral de potencia de un proceso estacionario.

Propiedad 1

El valor de frecuencia cero de la densidad espectral de potencia de un proceso estacionario es igual al área total bajo la gráfica de la función de autocorrelación; esto es,

$$S_X(0) = \int_{-\infty}^{\infty} R_X(\tau)\, d\tau \tag{1.44}$$

Esta propiedad sigue directamente de la ecuación (1.42) al hacer $f = 0$.

Propiedad 2

El valor cuadrático medio de un proceso estacionario es igual al área total bajo la gráfica de la densidad espectral de potencia; esto es,

$$E[X^2(t)] = \int_{-\infty}^{\infty} S_X(f)\, df \tag{1.45}$$

Esta propiedad sigue directamente de la ecuación (1.43) al sustituir $\tau = 0$ y advertir que $R_X(0) = E[X^2(t)]$.

Propiedad 3

La densidad espectral de potencia de un proceso estacionario siempre es no negativa; esto es,

$$S_X(f) \geq 0 \qquad \text{para toda } f \tag{1.46}$$

Esta propiedad es una consecuencia inmediata de que en la ecuación (1.41) el valor cuadrático medio $E[Y^2(t)]$ siempre debe ser no negativo.

Propiedad 4

La densidad espectral de potencia de un proceso aleatorio de valores reales es una función par de la frecuencia; esto es,

$$S_X(-f) = S_X(f) \tag{1.47}$$

Esta propiedad se obtiene al sustituir $-f$ por f en la ecuación (1.42):

$$S_X(-f) = \int_{-\infty}^{\infty} R_X(\tau) \exp(j2\pi f\tau)\, d\tau$$

En seguida, al sustituir $-\tau$ por τ y reconociendo que $R_X(-\tau) = R_X(\tau)$, obtenemos

$$S_X(-f) = \int_{-\infty}^{\infty} R_X(\tau) \exp(-j2\pi f\tau)\, d\tau = S_X(f)$$

que es el resultado deseado.

Propiedad 5

La densidad espectral de potencia, normalizada en forma apropiada, tiene las propiedades que suelen asociarse con una función de densidad de probabilidad.

La normalización que tenemos en mente aquí es con respecto al área total bajo la gráfica de la densidad espectral de potencia (es decir, el valor cuadrático medio del proceso). Considere en este caso la función

$$p_X(f) = \frac{S_X(f)}{\int_{-\infty}^{\infty} S_X(f)\, df} \tag{1.48}$$

A la luz de las propiedades 2 y 3, advertimos que $p_X(f) \geq 0$ para toda f. Además, el área total bajo la función $p_X(f)$ es la unidad. Por consiguiente, la forma normalizada de la densidad espectral de potencia, según se define en la ecuación (1.48), se comporta de manera similar a una función de densidad de probabilidad.

EJEMPLO 1.5 Onda senoidal con fase aleatoria (continuación)

Considere el proceso aleatorio $X(t) = A\cos(2\pi f_c t + \Theta)$, donde Θ es una variable aleatoria distribuida uniformemente sobre el intervalo $[-\pi, \pi]$. La función de autocorrelación de este proceso aleatorio está dada por la ecuación (1.17), que se reproduce aquí por conveniencia:

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau)$$

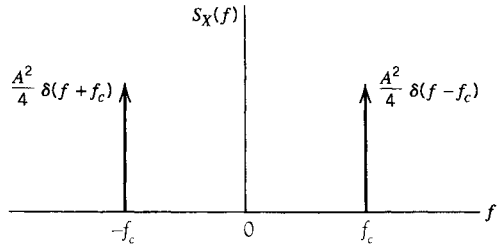


FIGURA 1.10 Densidad espectral de potencia de onda seno con fase aleatoria; $\delta(f)$ denota la función delta en $f = 0$.

Dejemos que $\delta(f)$ denote la *función delta* en $f = 0$; para la definición delta y sus propiedades, consulte el apéndice 2. Tomando la transformada de Fourier en ambos miembros de la relación que define a $R_X(\tau)$, encontramos que la densidad espectral de potencia del proceso senoidal $X(t)$ es

$$S_X(f) = \frac{A^2}{4}[\delta(f - f_c) + \delta(f + f_c)] \tag{1.49}$$

la cual consta de un par de funciones delta ponderadas por el factor $A^2/4$ y ubicadas en $\pm f_c$, como se ilustra en la figura 1.10. Notamos que el área total bajo la función delta es igual a uno. Por consiguiente, el área total bajo $S_X(f)$ es igual a $A^2/2$, como se esperaba. ◀

▶ EJEMPLO 1.6 Onda binaria aleatoria (continuación)

Considere de nuevo una onda binaria aleatoria compuesta de una secuencia de unos y ceros representada mediante los valores de $+A$ y $-A$, respectivamente. En el ejemplo 1.3 demostramos que la función de autocorrelación de este proceso aleatorio tiene una forma de onda triangular, como se indica mediante

$$R_X(\tau) = \begin{cases} A^2\left(1 - \dfrac{|\tau|}{T}\right), & |\tau| < T \\ 0, & |\tau| \geq T \end{cases}$$

La densidad espectral de potencia del proceso es consecuentemente

$$S_X(f) = \int_{-T}^{T} A^2\left(1 - \frac{|\tau|}{T}\right)\exp(-j2\pi f\tau)\, d\tau$$

Utilizando la transformada de Fourier de una función triangular (vea la tabla A6.3), obtenemos

$$S_X(f) = A^2 T \operatorname{senc}^2(fT) \tag{1.50}$$

la cual se grafica en la figura 1.11. Aquí también vemos que la densidad espectral de potencia es no negativa para toda f y que es una función par de f. Al advertir que $R_X(0) = A^2$ y al utilizar la propiedad 2, encontramos que el área total bajo $S_X(f)$, o la potencia promedio de la onda binaria aleatoria descrita aquí, corresponde a A^2, lo cual es intuitivamente satisfactorio. ◀

El resultado de la ecuación (1.50) se generaliza del modo siguiente. Advertimos que la *densidad espectral de energía* (es decir, la magnitud al cuadrado de la transformada de Fourier) de un pulso rectangular $g(t)$ de amplitud A y duración T está dada por

$$E_g(f) = A^2T^2\operatorname{senc}^2(fT) \tag{1.51}$$

Por lo tanto, es posible describir la ecuación (1.50) en términos de $E_g(f)$ simplemente como

$$S_X(f) = \frac{E_g(f)}{T} \tag{1.52}$$

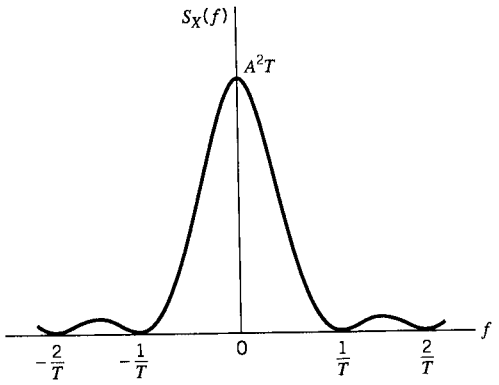


FIGURA 1.11 Densidad espectral de potencia de una onda binaria aleatoria.

La ecuación (1.52) establece que para una onda binaria aleatoria en la que los símbolos binarios 1 y 0 se representan por medio de los pulsos $g(t)$ y $-g(t)$, respectivamente, la densidad espectral de potencia $S_X(f)$ es igual a la densidad espectral de energía $E_g(f)$ del *pulso de formación del símbolo* $g(t)$, dividido entre la *duración del símbolo* T.

EJEMPLO 1.7 Mezcla de un proceso aleatorio con un proceso senoidal

Una situación que surge a menudo en la práctica es la de *mezcla* (es decir, multiplicación) de un proceso estacionario $X(t)$ como una onda senoidal $\cos(2\pi f_c t + \Theta)$, donde la fase Θ es una variable aleatoria que está distribuida de manera uniforme sobre el intervalo $[0, 2\pi]$. La adición de este modo de la fase aleatoria Θ tan sólo reconoce el hecho de que el origen de tiempo se elige arbitrariamente cuando $X(t)$ y $\cos(2\pi f_c t + \Theta)$ provienen de fuentes independientes físicamente, como suele ser el caso. Nos interesa determinar la densidad espectral de potencia del proceso aleatorio $Y(t)$, definido por

$$Y(t) = X(t)\cos(2\pi f_c t + \Theta) \tag{1.53}$$

Al utilizar la definición de la función de autocorrelación de un proceso estacionario y advertir que la variable aleatoria Θ es independiente de $X(t)$, encontramos que la función de autocorrelación de $Y(t)$ está dada por

$$\begin{aligned} R_Y(\tau) &= E[Y(t+\tau)Y(t)] \\ &= E[X(t+\tau)\cos(2\pi f_c t + 2\pi f_c \tau + \Theta)X(t)\cos(2\pi f_c t + \Theta)] \\ &= E[X(t+\tau)X(t)]E[\cos(2\pi f_c t + 2\pi f_c \tau + \Theta)\cos(2\pi f_c t + \Theta)] \\ &= \tfrac{1}{2}R_X(\tau)E[\cos(2\pi f_c \tau) + \cos(4\pi f_c t + 2\pi f_c \tau + 2\Theta)] \\ &= \tfrac{1}{2}R_X(\tau)\cos(2\pi f_c \tau) \end{aligned} \tag{1.54}$$

Debido a que la densidad espectral de potencia es la transformada de Fourier de la función de autocorrelación, encontramos que las densidades espectrales de potencia de los procesos aleatorios $X(t)$ y $Y(t)$ se relacionan del modo siguiente:

$$S_Y(f) = \tfrac{1}{4}[S_X(f - f_c) + S_X(f + f_c)] \tag{1.55}$$

De acuerdo con la ecuación (1.55), la densidad espectral de potencia del proceso aleatorio $Y(t)$ definido en la ecuación (1.53) se obtiene como sigue: desplazamos la densidad espectral de potencia dada $S_X(f)$ del proceso aleatorio $X(t)$ hacia la derecha por f_c, la desplazamos hacia la izquierda por f_c, sumamos los dos espectros de potencia desplazados y dividimos el resultado entre cuatro. ◀

■ RELACIÓN ENTRE LAS DENSIDADES ESPECTRALES DE POTENCIA DE PROCESOS ALEATORIOS DE ENTRADA Y SALIDA

Dejemos que $S_X(f)$ denote la densidad espectral de potencia del proceso aleatorio de salida $Y(t)$ obtenida al hacer pasar el proceso aleatorio $X(t)$ por un filtro lineal de respuesta en frecuencia $H(f)$. En ese caso, al reconocer por definición que la densidad espectral de potencia de un proceso aleatorio es igual a la transformada de Fourier de su función de autocorrelación y al utilizar la ecuación (1.32), obtenemos

$$\begin{aligned} S_Y(f) &= \int_{-\infty}^{\infty} R_Y(\tau) \exp(-j2\pi f\tau)\, d\tau \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_X(\tau-\tau_1+\tau_2)\exp(-j2\pi f\tau)\, d\tau_1\, d\tau_2 d\tau \end{aligned} \tag{1.56}$$

Sea $\tau - \tau_1 + \tau_2 = \tau_0$ o, equivalentemente $\tau = \tau_0 + \tau_1 - \tau_2$. Al hacer entonces esta sustitución en la ecuación (1.56), descubrimos que $S_X(f)$ se puede expresar como el producto de tres términos: la respuesta en frecuencia $H(f)$ del filtro, la compleja conjugada de $H(f)$ y la densidad espectral de potencia de $S_X(f)$ del proceso aleatorio de entrada $X(t)$. En consecuencia, es posible simplificar la ecuación (1.56):

$$S_Y(f) = H(f)H^*(f)S_X(f) \tag{1.57}$$

Por último, puesto que $|H(f)|^2 = H(f)H^*(f)$, encontramos que la relación entre las densidades espectrales de potencia de los procesos aleatorios de entrada y salida se expresa en el dominio de la frecuencia al escribir

$$S_Y(f) = |H(f)|^2 S_X(f) \tag{1.58}$$

La ecuación (1.58) establece que *la densidad espectral de potencia del proceso de salida $Y(t)$ es igual a la densidad de potencia del proceso de entrada $X(t)$ multiplicado por la respuesta en magnitud al cuadrado del filtro.* Mediante el uso de esta relación, es posible determinar consecuentemente el efecto de hacer pasar el proceso aleatorio por un filtro estable y lineal invariablemente. En términos computacionales, la ecuación (1.58) suele ser más fácil de manejar que su contraparte en el dominio del tiempo de la ecuación (1.32), incluyendo la función de autocorrelación.

■ RELACIÓN ENTRE LA DENSIDAD ESPECTRAL DE POTENCIA Y EL ESPECTRO EN MAGNITUD DE UNA FUNCIÓN MUESTRA

Deseamos relacionar la densidad espectral de potencia $S_X(f)$ directamente con las propiedades espectrales de una función muestra $x(t)$ de un proceso estacionario $X(t)$ que es ergódico. Sin embargo, para que la función muestra $x(t)$ sea transformada de Fourier, ésta debe ser absolutamente integrable; es decir,

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty \tag{1.59}$$

Esta condición nunca puede satisfacerla ninguna función muestra estacionaria $x(t)$ de duración infinita. Con el fin de utilizar la técnica de la transformada de Fourier, consideramos un segmento truncado de $x(t)$, definido sobre el intervalo observacional $-T \le t \le T$, por ejemplo. De tal modo, al utilizar $X(f, T)$ para denotar la transformada de Fourier de la función muestra truncada definida de ese modo, es posible escribir

$$X(f,T) = \int_{-T}^{T} x(t)\exp(-j2\pi ft)\, dt \tag{1.60}$$

Suponiendo que el proceso $x(t)$ también es ergódico, es factible evaluar la función de autocorrelación $R_X(\tau)$ de $X(t)$ utilizando la fórmula promediada en el tiempo (vea la sección 1.5)

$$R_X(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)\,dt \tag{1.61}$$

Suele considerarse la función muestra $x(t)$ como una *señal de potencia* (es decir, una señal con potencia promedio finita). Por tanto, podemos formular el siguiente par de transformadas de Fourier:

$$\frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)\,dt \rightleftharpoons \frac{1}{2T}|X(f,T)|^2 \tag{1.62}$$

El parámetro en el primer miembro es una función de autocorrelación promediada en el tiempo. El parámetro del segundo miembro recibe el nombre de *periodograma*, cuyas dimensiones son las mismas que las de la densidad espectral de potencia. Sin embargo, esta terminología es una designación errónea, ya que el periodograma es una función de la frecuencia, no del periodo. A pesar de eso, se usa ampliamente. La cantidad la utilizaron primero los estadísticos para tener en cuenta periodicidades, como las tendencias estacionales en los datos.

Al utilizar la fórmula para la transformada inversa de Fourier en el par de transformadas de Fourier en la ecuación (1.62), es posible expresar la función de autocorrelación promediada en el tiempo de la función muestra $x(t)$ en términos del periodograma como

$$\frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)\,dt = \int_{-\infty}^{\infty} \frac{1}{2T}|X(f,T)|^2 \exp(j2\pi fT)\,d \tag{1.63}$$

De ese modo, al sustituir la ecuación (1.63) en la (1.61), obtenemos

$$R_X(\tau) = \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{1}{2T}|X(f,T)|^2 \exp(j2\pi f\tau)\,df \tag{1.64}$$

Para un valor fijo de la frecuencia f, el periodograma es una variable aleatoria porque su valor varía de manera aleatoria de una función muestra del proceso aleatorio a otra. De esa manera, para una función muestra determinada $x(t)$, el periodograma no converge en ningún sentido estadístico hasta un valor límite cuando T tiende a infinito. De ese modo, resultaría incorrecto intercambiar el orden de las operaciones de integración y límite en la ecuación (1.64). No obstante, supongamos que tomamos la esperanza en ambos miembros de la ecuación (1.64) sobre el conjunto de todas las funciones muestra sobre el proceso aleatorio y reconocemos que en un proceso ergódico la función de autocorrelación $R_X(\tau)$ permanece invariable ante esta operación. En ese caso, puesto que cada función muestra de un proceso ergódico a la larga toma casi todos los modos de comportamiento de cada una de las otras funciones muestra , es posible escribir

$$R_X(\tau) = \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{1}{2T} E\left[|X(f,T)|^2\right] \exp(j2\pi f\tau)\,df \tag{1.65}$$

Enseguida podemos intercambiar el orden de integración y las operaciones límite para obtener

$$R_X(\tau) = \int_{-\infty}^{\infty} \left\{\lim_{T\to\infty} \frac{1}{2T} E\left[|X(f,T)|^2\right]\right\} \exp(j2\pi f\tau)\,df \tag{1.66}$$

Por consiguiente, al comparar las ecuaciones (1.66) y (1.43), obtenemos la relación deseada entre la densidad espectral de potencia $S_X(f)$ de un proceso ergódico y el espectro de magnitud al cuadrado $|X(f, T)|^2$ de una función muestra truncada del proceso:

$$\begin{aligned} S_X(f) &= \lim_{T\to\infty} \frac{1}{2T} E\left[|X(f,T)|^2\right] \\ &= \lim_{T\to\infty} \frac{1}{2T} E\left[\left|\int_{-T}^{T} x(t)\exp(-j2\pi ft)\,dt\right|^2\right] \end{aligned} \tag{1.67}$$

Es importante señalar que en la ecuación (1.67) no es posible hacer que $T \to \infty$ antes de tomar la esperanza. La ecuación (1.67) proporciona la base matemática para *estimar*[4] la densidad espectral de potencia de un proceso aleatorio ergódico, dada una función muestra $x(t)$ de un proceso observado sobre el intervalo $[-T, T]$.

■ DENSIDADES ESPECTRALES CRUZADAS

Del mismo modo que la densidad espectral de potencia proporciona una medida de la distribución de frecuencia de un proceso aleatorio simple, la densidad espectral cruzada ofrece una medida de la interrelación de frecuencia entre dos procesos aleatorios. En particular, sean $X(t)$ y $Y(t)$ dos procesos estacionarios conjuntos con sus funciones de correlación cruzada denotadas por $R_{XY}(\tau)$ y $R_{YX}(\tau)$. Definimos entonces las *densidades espectrales cruzadas* $S_{XY}(f)$ y $S_{YX}(f)$ de este par de procesos aleatorios como las transformadas de Fourier de sus funciones de correlación cruzada respectivas, según se indica por medio de

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.68}$$

y

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(\tau) \exp(-j2\pi f\tau)\, d\tau \tag{1.69}$$

Las funciones de correlación cruzada y las densidades espectrales cruzadas forman consecuentemente pares de transformadas de Fourier. Por tanto, utilizando la fórmula para la transformada inversa de Fourier también es posible escribir

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(f) \exp(j2\pi f\tau)\, df \tag{1.70}$$

y

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} S_{YX}(f) \exp(j2\pi f\tau)\, df \tag{1.71}$$

Las densidades espectrales cruzadas $S_{XY}(f)$ y $S_{YX}(f)$ no son necesariamente funciones reales de la frecuencia f. Sin embargo, al sustituir la relación

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

en la ecuación (1.68) y empleando después la ecuación (1.69) encontramos que $S_{XY}(f)$ y $S_{YX}(f)$ se relacionan por medio de

$$S_{XY}(f) = S_{YX}(-f) = S_{YX}^{*}(f) \tag{1.72}$$

▶ EJEMPLO 1.8

Suponga que los procesos aleatorios $X(t)$ y $Y(t)$ tienen media cero y son individualmente estacionarios. Considere el proceso aleatorio de suma

$$Z(t) = X(t) + Y(t)$$

El problema es determinar la densidad espectral de potencia de $Z(t)$.

La función de autocorrelación de $Z(t)$ está dada por

$$\begin{aligned} R_Z(t,u) &= E[Z(t)Z(u)] \\ &= E[(X(t)+Y(t))(X(u)+Y(u))] \\ &= E[X(t)X(u)]+E[X(t)Y(u)]+E[Y(t)X(u)]+E[Y(t)Y(u)] \\ &= R_X(t,u)+R_{XY}(t,u)+R_{YX}(t,u)+R_Y(t,u) \end{aligned}$$

Definiendo $\tau = t - u$, podemos escribir

$$R_Z(\tau) = R_X(\tau)+R_{XY}(\tau)+R_{YX}(\tau)+R_Y(\tau) \tag{1.73}$$

cuando los procesos estacionarios $X(t)$ y $Y(t)$ son también estacionarios conjuntamente. Por consiguiente, al tomar la transformada de Fourier en la ecuación (1.73), obtenemos

$$S_Z(f) = S_X(f)+S_{XY}(f)+S_{YX}(f)+S_Y(f) \tag{1.74}$$

De tal modo, vemos que las densidades espectrales cruzadas $S_{XY}(f)$ y $S_{YX}(f)$ representan los componentes espectrales que deben agregarse a las densidades espectrales de potencia individuales de un par de procesos aleatorios correlacionados para obtener la densidad espectral de potencia de su suma.

Cuando los procesos estacionarios $X(t)$ y $Y(t)$ no están correlacionados, las densidades espectrales cruzadas $S_{XY}(f)$ y $S_{YX}(f)$ no son cero, y por ello la ecuación (1.74) se reduce a

$$S_Z(f) = S_X(f)+S_Y(f) \tag{1.75}$$

Es posible generalizar este último resultado estableciendo que cuando existe multiplicidad de procesos estacionarios de media cero que no están correlacionados entre sí, la densidad espectral de potencia de su suma es igual a la suma de sus densidades espectrales de potencia individuales. ◀

▶ EJEMPLO 1.9

Considere el siguiente problema de hacer pasar dos procesos estacionarios conjuntamente por un par de filtros lineales, estables e independientes invariantes con el tiempo, como se muestra en la figura 1.12. En particular, suponga que el proceso aleatorio $X(t)$ es la entrada para el filtro de respuesta al impulso $h_1(t)$ y que el proceso aleatorio $Y(t)$ es la entrada para el filtro de respuesta al impulso $h_2(t)$. Deje que $V(t)$ y $Z(t)$ denoten los procesos aleatorios en las salidas respectivas de los filtros. La función de correlación cruzada de $V(t)$ y de $Z(t)$ es, en consecuencia,

$$\begin{aligned} R_{VZ}(t,u) &= E[V(t)Z(u)] \\ &= E\left[\int_{-\infty}^{\infty} h_1(\tau_1)X(t-\tau_1)\,d\tau_1 \int_{-\infty}^{\infty} h_2(\tau_2)Y(u-\tau_2)\,d\tau_2\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)E[X(t-\tau_1)Y(u-\tau_2)]\,d\tau_1\,d\tau_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)R_{XY}(t-\tau_1,u-\tau_2)\,d\tau_1\,d\tau_2 \end{aligned} \tag{1.76}$$

donde $R_{XY}(t, u)$ es la función de correlación cruzada de $X(t)$ y $Y(t)$. Debido a que los procesos aleatorios de entrada son estacionarios conjuntamente (por hipótesis), es posible dejar $\tau = t - u$ y por ello reescribir la ecuación (1.76) de la manera siguiente:

$$R_{VZ}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(\tau_1)h_2(\tau_2)R_{XY}(\tau-\tau_1+\tau_2)\,d\tau_1\,d\tau_2 \tag{1.77}$$

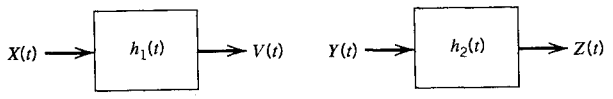


FIGURA 1.12 Un par de filtros lineales invariantes con el tiempo independientes.

Al tomar la transformada de Fourier en la ecuación (1.77) y recurrir a un procedimiento similar al que condujo a la ecuación (1.39), obtenemos finalmente

$$S_{VZ}(f) = H_1(f)H_2^*(f)S_{XY}(f) \tag{1.78}$$

donde $H_1(f)$ y $H_2(f)$ son las respuestas en frecuencia de los filtros respectivos en la figura 1.12, y $H_2^*(f)$ es la conjugada compleja de $H_2(f)$. Ésta es la relación deseada entre la densidad espectral cruzada de los procesos de salida y la de los procesos de entrada. ◀

1.8 *Proceso gaussiano*

El material que hemos presentado acerca de procesos aleatorios hasta este punto en la exposición ha sido de una naturaleza bastante general. En esta sección, consideraremos una importante familia de procesos aleatorios conocida como procesos gaussianos.[5]

Vamos a suponer que observamos un proceso aleatorio $X(t)$ en un intervalo que se inicia en el tiempo $t = 0$ y perdura hasta $t = T$. Supongamos también que ponderamos el proceso aleatorio $X(t)$ por medio de alguna función $g(t)$ y que luego integramos el producto $g(t)X(t)$ a lo largo de este intervalo de observación, obteniendo de esa manera una variable aleatoria Y definida por

$$Y = \int_0^T g(t)X(t)\,dt \tag{1.79}$$

Nos referimos a Y como una *funcional lineal* de $X(t)$. La distinción entre una función y una funcional debe señalarse con todo cuidado. Por ejemplo, la suma de $Y = \sum_{i=1}^{N} a_i X_i$, donde las a_i son constantes y las X_i son variables aleatorias, es una *función* lineal de las X_i; para cada conjunto observado de valores correspondientes a las variables aleatorias X_i, tenemos un valor correspondiente para la variable aleatoria Y. Por otro lado, en la ecuación (1.79) el valor de la variable aleatoria Y depende del curso de la *función argumento* $g(t)X(t)$ sobre el intervalo de observación completo de 0 a T. De tal modo, una funcional es una cantidad que depende del uso completo de una o más funciones en vez de un número de variables discretas. En otras palabras, el dominio de una funcional es un conjunto o espacio de funciones admisibles en vez de una región de un espacio de coordenadas.

Si en la ecuación (1.79) la función ponderadora $g(t)$ es tal que el valor cuadrático medio de la variable aleatoria Y es finito, y si Y es una variable aleatoria que se *distribuye gaussianamente* para toda $g(t)$ en esta clase de funciones, entonces se dice que el proceso $X(t)$ es un *proceso gaussiano*. En otras palabras, el proceso $X(t)$ es gaussiano si cada funcional lineal de $X(t)$ es una variable aleatoria gaussiana.

Decimos que la variable aleatoria Y tiene una distribución gaussiana si su función de densidad de probabilidad tiene la forma

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] \tag{1.80}$$

donde μ_Y es la media y σ_Y^2 es la varianza de la variable aleatoria Y. Una gráfica de esta función de densidad de probabilidad se presenta en la figura 1.13 para el caso especial en el que la variable aleatoria gaussiana Y se *normaliza* para tener una media μ_Y de cero y una varianza σ_Y^2 de uno, de acuerdo con

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)$$

Tal distribución gaussiana normalizada se escribe por lo común como $\mathcal{N}(0, 1)$.

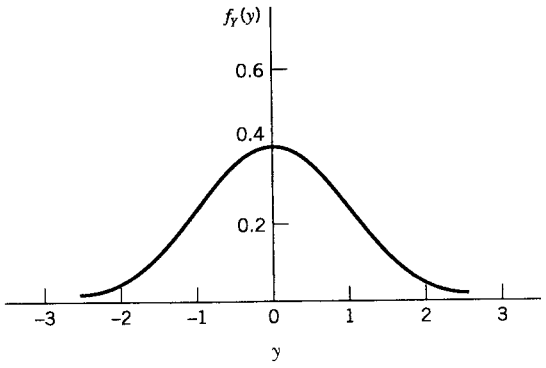


FIGURA 1.13 Distribución gaussiana normalizada.

Un proceso gaussiano tiene dos virtudes. Primero, el proceso gaussiano posee muchas propiedades que hacen posibles los resultados analíticos; analizaremos estas propiedades más adelante en la sección. Segundo, los procesos aleatorios producidos por fenómenos físicos son tales que a menudo resulta apropiado un modelo gaussiano. Además, el uso de un modelo gaussiano para describir los fenómenos físicos suele confirmarse mediante experimentos. De tal manera, la ocurrencia frecuente de fenómenos físicos para los cuales es apropiado el modelo gaussiano, junto con la facilidad con la que el proceso gaussiano se maneja matemáticamente, hacen que este último sea muy importante en el estudio de los sistemas de comunicación.

TEOREMA DEL LÍMITE CENTRAL

El *teorema del límite central* proporciona la justificación matemática para utilizar un proceso gaussiano como un modelo en un gran número de fenómenos físicos diferentes en los que la variable aleatoria observada, en un instante particular, es el resultado de un número grande de eventos aleatorios individuales. Para formular este importante teorema, sea X_i, $i = 1, 2, \ldots, N$, un conjunto de variables aleatorias que satisfacen los siguientes requerimientos:

1. Las X_i son independientes estadísticamente.
2. Las X_i tienen la misma distribución de probabilidad con media μ_X y varianza σ_X^2.

Las X_i descritas de este modo se dice que constituyen un conjunto de variables aleatorias *distribuidas independiente e idénticamente* (dii). Considérese que estas variables aleatorias se *normalizan* como sigue:

$$Y_i = \frac{1}{\sigma_X}(X_i - \mu_X), \quad i = 1, 2, \ldots, N$$

de modo que tenemos

$$E[Y_i] = 0$$

y

$$\text{var}[Y_i] = 1$$

Defina la variable aleatoria

$$V_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i$$

El teorema del límite central establece que la distribución de probabilidad de V_N *se aproxima a la distribución gaussiana normalizada* $\mathcal{N}(0, 1)$ *en el límite cuando el número de variables aleatorias* N *tiende a infinito.*

Sin embargo, es importante darse cuenta de que el teorema del límite central sólo proporciona la forma "límite" de la distribución de probabilidad de la variable aleatoria normalizada V_N cuando N tiende a infinito. Cuando N es finita, se encuentra algunas veces que el límite gaussiano ofrece una aproximación relativamente pobre para la distribución de probabilidad real de V_N aun cuando N quizá sea bastante grande.

■ PROPIEDADES DE UN PROCESO GAUSSIANO

Un proceso gaussiano tiene algunas propiedades útiles que se describen a continuación.

Propiedad 1

Si se aplica un proceso gaussiano $X(t)$ *a un filtro lineal estable, entonces el proceso aleatorio* $Y(t)$ *que se desarrolla a la salida del filtro también es gaussiano.*

Esta propiedad se deduce fácilmente utilizando la definición de un proceso gaussiano a partir de la ecuación (1.79). Considere la situación que se describe en la figura 1.8, donde tenemos un filtro invariante con el tiempo de respuesta al impulso $h(t)$, con el proceso aleatorio $X(t)$ como entrada y el proceso aleatorio $Y(t)$ como salida. Suponemos que $X(t)$ es un proceso gaussiano. Los procesos aleatorios $Y(t)$ y $X(t)$ se relacionan por medio de la integral de convolución

$$Y(t) = \int_0^T h(t-\tau)X(\tau)\,d\tau, \quad 0 \le t < \infty \tag{1.81}$$

Suponemos que la respuesta al impulso $h(t)$ es tal que el valor cuadrático medio del proceso aleatorio de salida $Y(t)$ es finito para toda t en el intervalo $0 \le t < \infty$ para el cual se define $Y(t)$. Para demostrar que el proceso de salida $Y(t)$ es gaussiano, debemos demostrar que cualquier función lineal de éste es una variable aleatoria gaussiana. Esto es, si definimos la variable aleatoria

$$Z = \int_0^\infty g_Y(t)\int_0^T h(t-\tau)X(\tau)\,d\tau\,dt \tag{1.82}$$

entonces Z debe ser una variable aleatoria gaussiana para toda función $g_Y(t)$, tal que el valor cuadrático medio de Z sea finito. Al intercambiar el orden de integración en la ecuación (1.82), obtenemos

$$Z = \int_0^T g(\tau)X(\tau)\,d\tau \tag{1.83}$$

donde

$$g(\tau) = \int_0^\infty g_Y(t)h(t-\tau)\,d\tau \tag{1.84}$$

Puesto que $X(t)$ es un proceso gaussiano por hipótesis, se concluye de la ecuación (1.83) que Z debe ser una variable aleatoria gaussiana. Así hemos demostrado que si la entrada $X(t)$ para un filtro lineal es un proceso gaussiano, entonces la salida $Y(t)$ también es un proceso gaussiano. Sin embargo, advierta que a pesar de que la demostración se realizó suponiendo un filtro lineal invariante con el tiempo, esta propiedad es válida para cualquier sistema lineal estable arbitrario.

Propiedad 2

Considere el conjunto de variables o muestras aleatorias $X(t_1), X(t_2), \ldots, X(t_n)$, *obtenido al observar un proceso aleatorio* $X(t)$ *en los tiempos* $t_1, t_2, \ldots, t_n$. *Si el proceso* $X(t)$ *es gaussiano, entonces este conjunto de variables aleatorias es gaussiano conjuntamente para toda n, estando su función de densidad de probabilidad conjunta enésima determinada completamente al especificar el conjunto de medias*

$$\mu_{X(t_i)} = E[X(t_i)], \quad i = 1, 2, \ldots, n$$

y *el conjunto de funciones de covarianza*

$$C_X(t_k, t_i) = E\left[(X(t_k) - \mu_{X(t_k)})(X(t_i) - \mu_{X(t_i)})\right], \quad k, i = 1, 2, \ldots, n$$

Consideremos que el vector $\mathbf{X}$ de n por 1 denota el conjunto de variables aleatorias $X(t_1), \ldots, X(t_n)$ derivado del proceso gaussiano $X(t)$ muestreándolo en los tiempos $t_1, \ldots, t_n$. Déjese que $\mathbf{x}$ denote el valor de $\mathbf{X}$. De acuerdo con la propiedad 2, el vector aleatorio $\mathbf{X}$ tiene una *distribución gaussiana multivariada* definida en forma de matriz como

$$f_{X(t_1),\ldots,X(t_2)}(x_1, \ldots, x_2) = \frac{1}{(2\pi)^{n/2}\Delta^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \tag{1.85}$$

donde el superíndice T denota transposición y

$\boldsymbol{\mu}$ = vector de la media

$\boldsymbol{\Sigma} = [\mu_1, \mu_2, \ldots, \mu_n]^T$

$\boldsymbol{\Sigma}$ = matriz de covarianza

$= \{C_X(t_k, t_i)\}_{k,i=1}^n$

$\boldsymbol{\Sigma}$ = inversa de la matriz de covarianza $\boldsymbol{\Sigma}$

$\boldsymbol{\Sigma}$ = determinante de la matriz de covarianza $\boldsymbol{\Sigma}$

La propiedad 2 se usa a menudo como la definición de un proceso gaussiano. Sin embargo, esta definición resulta más difícil de usar que la basada en la ecuación (1.79) para evaluar los efectos del filtrado en un proceso gaussiano.

Es posible extender la propiedad 2 a dos (o más) procesos aleatorios de la manera siguiente. Considere el conjunto compuesto de variables aleatorias $X(t_1), X(t_2), \ldots, X(t_n), Y(u_1), Y(u_2), \ldots, Y(u_m)$ obtenido al observar un proceso aleatorio $X(t)$ en los tiempos $\{t_i, i = 1, 2, \ldots, n\}$, y un segundo proceso aleatorio $Y(t)$ en los tiempos $\{u_k, k = 1, 2, \ldots, m\}$. Decimos que los procesos $X(t)$ y $Y(t)$ son *gaussianos conjuntamente* si este conjunto compuesto de variables aleatorias es gaussiano conjuntamente para toda n y m. Advierta que además de la media y las funciones de correlación de los procesos aleatorios $X(t)$ y $Y(t)$ por separado, también debemos conocer la función de covarianza cruzada

$$E\left[(X(t_i) - \mu_{X(t_i)})(Y(u_k) - \mu_{Y(u_k)})\right] = R_{XY}(t_i, u_k) - \mu_{X(t_i)}\mu_{Y(u_k)}$$

para cualquier par de instantes de observación (t_i, u_k). Este conocimiento adicional está incorporado en la función de correlación cruzada, $R_{XY}(t_i, u_k)$, de los dos procesos $X(t)$ y $Y(t)$.

Propiedad 3

Si un proceso gaussiano es estacionario, entonces el proceso también es estrictamente estacionario.

Esto se deduce directamente de la propiedad 2.

Propiedad 4

Si las variables aleatorias $X(t_1), X(t_2), \ldots, X(t_n)$, *obtenidas mediante un proceso gaussiano* $X(t)$ *en los tiempos* $t_1, t_2, \ldots, t_n$, *no están correlacionadas, esto es,*

$$E\left[(X(t_k)-\mu_{X(t_k)})(X(t_i)-\mu_{X(t_i)})\right]=0, \quad i \neq k$$

entonces estas variables aleatorias son independientes estadísticamente.

La falta de correlación de $X(t_1), \ldots, X(t_n)$ significa que la matriz de covarianza $\mathbf{\Sigma}$ es una matriz diagonal, como se indica por medio de

$$\mathbf{\Sigma}=\begin{bmatrix}\sigma_1^2 & & \mathrm{O}\\ & \sigma_2^2 & \\ & & \ddots \\ \mathrm{O} & & \sigma_n^2\end{bmatrix}$$

donde

$$\sigma_i^2 = E\left[(X(t_i)-E[X(t_i)])^2\right], \quad i=1,2,\ldots,n$$

Bajo esta condición, la distribución gaussiana multivariada de la ecuación (1.85) se simplifica a

$$f_{\mathbf{X}}(\mathbf{x})=\prod_{i=1}^{n} f_{X_i}(x_i)$$

donde $X_i = X(t_i)$ y

$$f_{X_i}(x_i)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x_i-\mu_{X_i})^2}{2\sigma_i^2}\right)$$

En palabras, si las variables aleatorias gaussianas $X(t_1), \ldots, X(t_n)$ no están correlacionadas, entonces son independientes estadísticamente, lo que a su vez significa que la función de densidad de probabilidad conjunta de este conjunto de variables aleatorias puede expresarse como el producto de las funciones de densidad de probabilidad de las variables aleatorias individuales en el conjunto.

1.9 *Ruido*

El término *ruido* suele usarse para designar señales no deseadas que tienden a perturbar la transmisión y el procesamiento de señales en sistemas de comunicación y sobre las cuales no tenemos un control completo. En la práctica, encontramos que existen muchas fuentes potenciales de ruido en un sistema de comunicación. Éstas pueden ser externas al sistema (por ejemplo, ruido atmosférico, ruido galáctico, ruido hecho por el hombre) o internas al sistema. La segunda categoría incluye un tipo importante de ruido que surge de *fluctuaciones espontáneas* de la corriente o voltaje en circuitos eléctricos.[6] Este tipo de ruido representa una limitación básica en la transmisión o detección de señales en sistemas de comunicación que incluyen el uso de dispositivos electrónicos. Los dos ejemplos más comunes de fluctuaciones espontáneas en circuitos eléctricos son el *ruido de disparo* y el *ruido térmico*, que se describen en seguida.

■ RUIDO DE DISPARO

El ruido de disparo surge en dispositivos electrónicos como los diodos y los transistores debido a la naturaleza discreta del flujo de corriente en estos dispositivos. Por ejemplo, en un circuito

fotodetector se genera un pulso de corriente cada vez que el cátodo emite un electrón debido al uso incidente de una fuente de intensidad constante. Los electrones se emiten naturalmente en tiempos aleatorios denotados por τ_k, donde $-\infty < k < \infty$. Se supone que las emisiones aleatorias de los electrones se han realizado durante largo tiempo. De ese modo, la corriente total que fluye por el fotodetector puede modelarse como una suma infinita de pulsos de corriente:

$$X(t) = \sum_{k=-\infty}^{\infty} h(t - \tau_k) \tag{1.86}$$

donde $h(t - \tau_k)$ es el pulso de corriente generado en el tiempo τ_k. El proceso $X(t)$definido por la ecuación (1.86) es uno estacionario denominado *ruido de disparo.*

El número de electrones $N(t)$, emitido en el intervalo de tiempo $[0, t]$ constituye un proceso estocástico discreto, el valor del cual aumenta en uno cada vez que se emite un electrón. La figura 1.14 ejemplifica una función muestra de este tipo. Dejemos que el valor medio del número de electrones, v, emitido entre los tiempos t y $t + t_0$ se defina por medio de

$$E[v] = \lambda t_0 \tag{1.87}$$

El parámetro λ es una constante denominada la *velocidad* del proceso. El número total de electrones emitido en el intervalo $[t, t + t_0]$, esto es,

$$v = N(t + t_0) - N(t)$$

sigue una *distribución de Poisson* con un valor medio igual a λt_0. En particular, la probabilidad de que k electrones emita en el intervalo $[t, t + t_0]$ está definida por

$$P(v = k) = \frac{(\lambda t_0)^k}{k!} e^{-\lambda t_0} \quad k = 0, 1, \ldots \tag{1.88}$$

Desafortunadamente, una detallada caracterización estadística del proceso de ruido de disparo $X(t)$ definido en la ecuación (1.86) resulta una difícil tarea matemática. Aquí simplemente señalaremos los resultados relativos a los primeros dos momentos del proceso:

- La media de $X(t)$ es

$$\mu_X = \lambda \int_{-\infty}^{\infty} h(t)\, dt \tag{1.89}$$

donde λ es la velocidad del proceso y $h(t)$ es la velocidad de onda de un pulso de corriente.

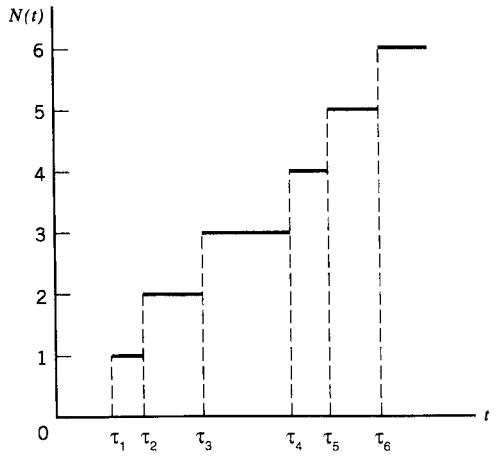


FIGURA 1.14 Función muestra de un proceso de conteo de Poisson.

▶ La función de autocovarianza de $X(t)$ es

$$C_X(\tau) = \lambda \int_{-\infty}^{\infty} h(t)h(t+\tau)\, dt \tag{1.90}$$

Este segundo resultado se conoce como *teorema de Campbell.*

En el caso especial de una forma de onda $h(t)$ consistente en un pulso rectangular de amplitud A y duración T, la media de proceso de ruido de disparo $X(t)$ es λAT, y su función de autocovarianza es

$$C_X(\tau) = \begin{cases} \lambda A^2(T - |\tau|), & |\tau| < T \\ 0, & |\tau| \geq T \end{cases}$$

que tiene una forma triangular similar a la que se presenta en la figura 1.7.

■ RUIDO TÉRMICO

El *ruido térmico* es el nombre dado al ruido eléctrico que surge del movimiento aleatorio de los electrones en un conductor. El valor medio cuadrático del voltaje del ruido térmico V_{TN} que aparece en las terminales de un resistor, medido en un ancho de banda Δf en Hertz, está, para todos los fines prácticos, dado por

$$E[V_{TN}^2] = 4kTR\,\Delta f \text{ volts}^2 \tag{1.91}$$

donde k es la *constante de Boltzman* igual a 1.38×10^{-23} joules por grado Kelvin, T es la *temperatura absoluta* en grados Kelvin y R es la resistencia en ohms. De este modo, es posible que modelemos un resistor ruidoso por medio del *circuito equivalente de Thévenin* compuesto por un generador de voltaje de ruido de valor cuadrático medio $E[V_{TN}^2]$ en serie con un resistor sin ruido, como en la figura 1.15*a*. De manera alternativa, podemos recurrir al *circuito equivalente de Norton* compuesto por un generador de corriente de ruido en paralelo con una conductancia sin ruido, como en la figura 1.15*b*. El valor medio cuadrático del generador de corriente de ruido es

$$\begin{aligned} E[I_{TN}^2] &= \frac{1}{R^2} E[V_{TN}^2] \\ &= 4kTG\,\Delta f \text{ amps}^2 \end{aligned} \tag{1.92}$$

donde $G = 1/R$ es la conductancia. También resulta de interés notar que debido a que el número de electrones en un resistor es muy grande y a que sus movimientos aleatorios dentro del resistor son independientemente estadísticos entre sí, el teorema del límite central indica que el ruido térmico se distribuye en forma gaussiana con media igual a cero.

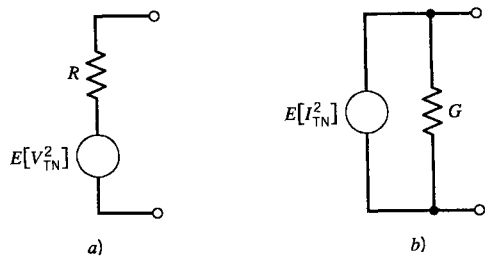


FIGURA 1.15 Modelos de un resistor ruidoso. *a*) Circuito equivalente de Thévenin. *b*) Circuito equivalente de Norton.

Los cálculos de ruido incluyen la transferencia de potencia, por lo que encontramos que el uso del *teorema de máxima transferencia de potencia* resulta aplicable a dichos cálculos. Este teorema establece que la máxima potencia posible se transfiere de una fuente de resistencia interna R a una carga de resistencia R_l cuando $R_l = R$. Bajo esta *condición de acoplamiento* la potencia que produce la fuente se divide igualmente entre la resistencia interna de la fuente y la resistencia de la carga, y la potencia entregada a la carga se denomina *potencia disponible.* Al aplicar el teorema de máxima transferencia de potencia al circuito equivalente de Thévenin de la figura 1.15*a* o al circuito equivalente de Norton de la figura 1.15*b*, encontramos que el resistor ruidoso produce una *potencia de ruido disponible* igual a $kT\,\Delta f$ watts.

Ruido blanco

El análisis de ruido de los sistemas de comunicación suele basarse en una forma idealizada de ruido denominada *ruido blanco,* cuya densidad espectral de potencia es independiente de la frecuencia de operación. El adjetivo *blanco* se usa porque la luz blanca contiene cantidades iguales de todas las frecuencias dentro de la banda visible de radiación electromagnética. Expresamos la densidad espectral de potencia del ruido blanco, con una función de potencia denotada por $w(t)$, como

$$S_W(f) = \frac{N_0}{2} \tag{1.93}$$

que se ilustra en la figura 1.16*a*. Las dimensiones de N_0 son watts por Hertz. El parámetro N_0 casi siempre se corresponde con la etapa de entrada del receptor de un sistema de comunicación. Es posible expresarlo como

$$N_0 = kT_e \tag{1.94}$$

donde k es la constante de Boltzmann y T_e la *temperatura de ruido equivalente* del receptor.[7] La temperatura de ruido equivalente de un sistema se define como *la temperatura a la cual un resistor ruidoso tiene que mantenerse de tal modo que, al conectar el resistor a la entrada de una versión sin ruido del sistema, produce la misma potencia de ruido disponible en la salida del sistema que la generada por todas las fuentes de ruido en el sistema real.* La característica importante de la temperatura del ruido equivalente es que depende sólo de los parámetros del sistema.

Puesto que la función de autocorrelación es la transformada inversa de Fourier de la densidad espectral de potencia, se concluye que para el ruido blanco

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau) \tag{1.95}$$

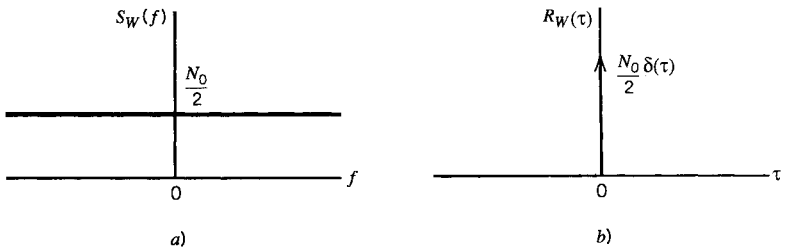


FIGURA 1.16 Características del ruido blanco. *a*) Densidad espectral de potencia. *b*) Función de autocorrelación.

Esto es, la función de autocorrelación del ruido blanco consiste en una función delta generada por el factor $N_0/2$ y que ocurre en $\tau = 0$, como en la figura 1.16*b*. Advertimos que $R_W(\tau)$ es cero para $\tau \neq 0$. En consecuencia, cualesquiera dos muestras diferentes de ruido blanco, no importa qué tan próximas entre sí se tomen en el tiempo, no están correlacionadas. Si el ruido blanco $w(t)$ también es gaussiano, entonces las dos muestras resultan estadísticamente independientes. En cierto sentido, el ruido blanco representa lo último en "aleatoriedad".

En términos generales, el ruido blanco tiene una potencia promedio infinita y, como tal, no puede ser físicamente realizable. A pesar de eso, el ruido blanco tiene propiedades matemáticas simples ejemplificadas por las ecuaciones (1.93) y (1.95), mismas que lo hacen útil en el análisis de sistemas estadísticos.

La utilidad del proceso de ruido blanco es paralela a la de la función impulso o de la función delta en el análisis de sistemas lineales. Del mismo modo que es posible observar el efecto de un impulso sólo después de que éste ha pasado por un sistema con un ancho de banda finito, con el ruido blanco ocurre igual, pues su efecto se observa sólo después de que pasa por un sistema similar. Por tanto, es posible establecer que siempre que el ancho de banda de un proceso de ruido en la entrada de un sistema es apreciablemente mayor que el del sistema mismo, entonces se puede modelar el proceso de ruido como ruido blanco.

EJEMPLO 1.10 Ruido blanco de filtrado pasobajas ideal

Suponga que un ruido gaussiano blanco $w(t)$ de media cero y densidad espectral de potencia $N_0/2$ se aplica a un filtro pasobajas ideal de un ancho de banda B y respuesta en magnitud pasobanda de uno. La densidad espectral de potencia del ruido $n(t)$ que aparece a la salida del filtro es, en consecuencia (vea la figura 1.17),

$$S_N(f) = \begin{cases} \dfrac{N_0}{2}, & -B < f < B \\ 0, & |f| > B \end{cases} \tag{1.96}$$

La función de autocorrelación de $n(t)$ es la transformada inversa de Fourier de la densidad espectral de potencia que se presenta en la figura 1.17*a*:

$$\begin{aligned} R_N(\tau) &= \int_{-B}^{B} \frac{N_0}{2} \exp(j2\pi f\tau)\, df \\ &= N_0 B \operatorname{senc}(2B\tau) \end{aligned} \tag{1.97}$$

Esta función de autocorrelación se grafica en la figura 1.17*b*. Vemos que $R_N(\tau)$ tiene su valor máximo de N_0B en el origen, y que pasa por cero en $\tau = \pm k/2B$, donde $k = 1, 2, 3, \ldots$.

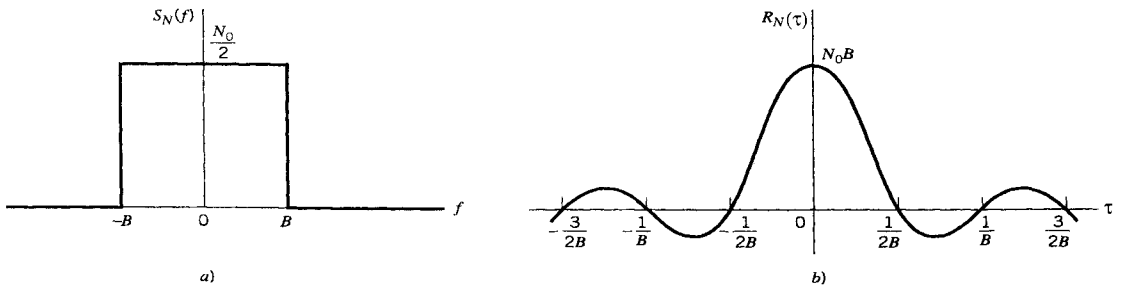


FIGURA 1.17 Características de ruido blanco del filtrado pasobajas. *a*) Densidad espectral de potencia. *b*) Función de autocorrelación.

Puesto que el ruido de entrada $w(t)$ es gaussiano (por hipótesis), se concluye que el ruido limitado en banda $n(t)$ en la salida del filtro también es gaussiano. Suponga que $n(t)$ se muestrea a una velocidad de $2B$ por segundo. De acuerdo con la figura 1.17*b*, vemos que las muestras de ruido resultantes no están correlacionadas y, al ser gaussianas, son estadísticamente independientes. Por consiguiente, la función de densidad de probabilidad conjunta de un conjunto de muestras de ruido obtenido de esta manera es igual al producto de las funciones de densidad de probabilidad individuales. Advierta que cada una de estas muestras de ruido tiene una media cero y una varianza de N_0B. ◀

▶ EJEMPLO 1.11 Correlación del ruido blanco con una onda senoidal

Considere la función muestra

$$w'(t) = \sqrt{\frac{2}{T}} \int_0^T w(t) \cos(2\pi f_c t)\, dt \tag{1.98}$$

que es la salida de un correlator con ruido gaussiano blanco $w(t)$ y una onda senoidal $\sqrt{2/T}\cos(2\pi f_c t)$ como entradas; el factor de escalamiento $\sqrt{2/T}$ se incluye aquí para lograr que la entrada de la onda senoidal tenga energía unitaria sobre el intervalo $0 \le t \le T$. (Este problema se encontró en el capítulo de antecedentes y conceptos preliminares, aunque no se desarrolló en esa ocasión.) Con el ruido $w(t)$ con media cero, se concluye de inmediato que la salida del correlator $w'(t)$ tiene también media cero. La varianza de la salida del correlator está definida por

$$\begin{aligned}
\sigma^2 &= E\left[\frac{2}{T}\int_0^T \int_0^T w(t_1)\cos(2\pi f_c t_1) w(t_2)\cos(2\pi f_c t_2)\, dt_1\, dt_2\right] \\
&= \frac{2}{T}\int_0^T \int_0^T E[w(t_1)w(t_2)]\cos(2\pi f_c t_1)\cos(2\pi f_c t_2)\, dt_1\, dt_2 \\
&= \frac{2}{T}\int_0^T \int_0^T R_W(t_1, t_2)\cos(2\pi f_c t_1)\cos(2\pi f_c t_2)\, dt_1\, dt_2
\end{aligned}$$

donde $R_W(t_1, t_2)$ es la función de autocorrelación del ruido blanco $w(t)$. Pero según la ecuación (1.95):

$$R_W(t_1, t_2) = \frac{N_0}{2}\delta(t_1 - t_2)$$

donde $N_0/2$ es la densidad espectral de potencia del ruido blanco $w(t)$. En consecuencia, es posible simplificar la expresión para la varianza σ^2 como

$$\sigma^2 = \frac{N_0}{2}\cdot\frac{2}{T}\int_0^T \int_0^T \delta(t_1 - t_2)\cos(2\pi f_c t_1)\cos(2\pi f_c t_2)\, dt_1\, dt_2$$

Apelamos ahora a la *propiedad de filtrado* de la función delta, a saber,

$$\int_{-\infty}^{\infty} g(t)\,\delta(t)\, dt = g(0)$$

donde $g(t)$ es una función continua del tiempo, suponiendo el valor $g(0)$ en el tiempo $t = 0$. Por tanto, aún puede simplificarse más σ^2:

$$\begin{aligned}
\sigma^2 &= \frac{N_0}{2}\cdot\frac{2}{T}\int_0^T \cos^2(2\pi f_c t)\, dt \\
&= \frac{N_0}{2}
\end{aligned} \tag{1.99}$$

donde se supone que la frecuencia f_c de la entrada de la onda senoidal es un múltiplo entero del recíproco de T. ◀

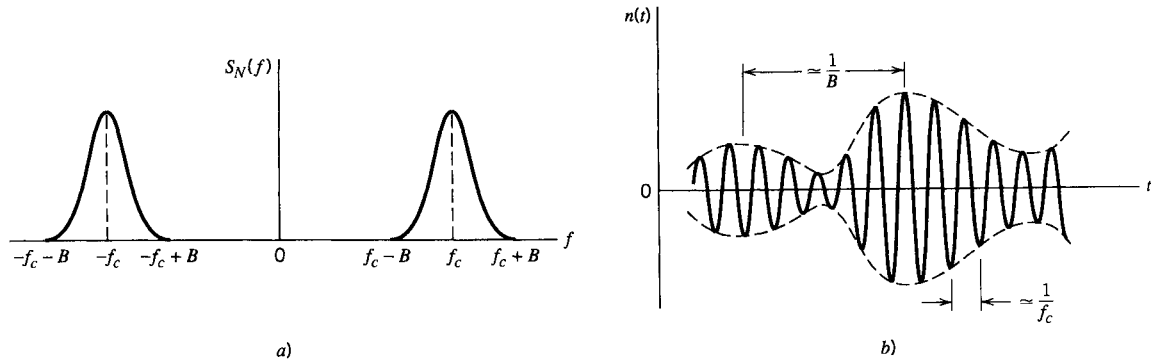


FIGURA 1.18 *a*) Densidad espectral de potencia de ruido de banda angosta. *b*) Función de muestreo de ruido de banda angosta.

1.10 *Ruido de banda angosta*

El receptor de un sistema de comunicación puede incluir alguna provisión para el *preprocesamiento* de la señal recibida. El preprocesamiento puede tomar la forma de un filtro de banda angosta cuyo ancho de banda es apenas lo suficientemente grande para dejar pasar la componente modulada de la señal recibida esencialmente no distorsionada, aunque no lo bastante grande para admitir ruido excesivo a través del receptor. El proceso de ruido que aparece en la salida de un filtro de estas características se conoce como *ruido de banda angosta.* Con las componentes espectrales del ruido de banda angosta concentradas en torno a cierta frecuencia de banda media $\pm f_c$ como en la figura 1.18*a*, encontramos que una función de muestreo $n(t)$ de tal proceso guarda cierta similitud con una onda seno de frecuencia f_c, que ondula lentamente tanto en amplitud como en fase, según se ilustra en la figura 1.18*b*.

Para analizar los efectos del ruido de banda angosta en el desempeño de un sistema de comunicación, necesitamos una representación matemática del mismo. Dependiendo de la aplicación de interés, existen dos representaciones específicas del ruido de banda angosta:

1. El ruido de banda angosta se define en términos de un par de componentes denominadas componente *en fase* y en *cuadratura.*
2. El ruido de banda angosta se define en términos de otras dos componentes denominadas la *envolvente* y la *fase.*

Estas dos representaciones se describen en lo que sigue. Por ahora resulta suficiente afirmar que dadas las componentes en fase y en cuadratura, es posible determinar las componentes de envolvente y de fase, y viceversa. Además, por propia cuenta, las dos representaciones no sólo son básicas para el análisis de ruido de los sistemas de comunicación, sino también para la caracterización del propio ruido de banda angosta.

1.11 *Representación del ruido de banda angosta en términos de las componentes en fase y en cuadratura*

Considere un ruido de banda angosta $n(t)$ de ancho de banda $2B$ centrado en la frecuencia f_c, como se ilustra en la figura 1.18. De acuerdo con la teoría de señales y sistemas pasobanda que se presenta en el apéndice 2, podemos representar $n(t)$ en la forma canónica (estándar):

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \tag{1.100}$$

donde $n_I(t)$ se denomina la *componente en fase* de $n(t)$, y $n_Q(t)$ se nombra como la *componente de cuadratura* de $n(t)$. Tanto $n_I(t)$ como $n_Q(t)$ son señales pasobajas. Salvo por la frecuencia de banda media f_c, estas dos componentes son por completo representativas del ruido de banda angosta $n(t)$.

Dado el ruido de banda angosta $n(t)$, es posible extraer sus componentes en fase y en cuadratura utilizando el esquema indicado en la figura 1.19*a*. Se supone que los dos filtros pasobajas que se utilizan en este esquema son ideales, cada uno con un ancho de banda igual a B (es decir, la mitad del ancho de banda del ruido de banda angosta $n(t)$). El esquema de la figura 1.19*a* sigue de la representación de la ecuación (1.100). Podemos, desde luego, utilizar directamente esta ecuación para generar el ruido de banda angosta $n(t)$, dadas sus componentes en fase y en cuadratura, como se muestra en la figura 1.19*b*. En consecuencia, los esquemas de las figuras 1.19*a* y 1.19*b* pueden considerarse como un *analizador* y un *sintetizador*, respectivamente, de ruido de banda angosta.

Los componentes en fase y en cuadratura de un ruido de banda angosta tienen propiedades importantes que se resumen a continuación:

1. El componente en fase $n_I(t)$ y el componente en cuadratura $n_Q(t)$ del ruido de banda angosta $n(t)$ tienen media cero.
2. Si el ruido de banda angosta $n(t)$ es gaussiano, entonces su componente en fase $n_I(t)$ y componente en cuadratura $n_Q(t)$ son gaussianas conjuntamente.
3. Si el ruido de banda angosta $n(t)$ es estacionario, entonces su componente en fase $n_I(t)$ y componente en cuadratura $n_Q(t)$ son estacionarias conjuntamente.
4. Tanto la componente en fase $n_I(t)$ como la componente en cuadratura $n_Q(t)$ tienen la misma densidad espectral de potencia, la cual se relaciona con la densidad espectral de potencia $S_N(f)$ del ruido de banda angosta $n(t)$ de acuerdo con

$$S_{N_I}(f) = S_{N_Q}(f) = \begin{cases} S_N(f - f_c) + S_N(f + f_c), & -B \le f \le B \\ 0, & \text{en otro caso} \end{cases} \tag{1.101}$$

   donde se supone que $S_N(f)$ ocupa el intervalo de frecuencia $f_c - B \le |f| \le f_c + B$ y $f_c > B$.
5. La componente en fase $n_I(t)$ y la componente en cuadratura $n_Q(t)$ tienen la misma varianza que el ruido de banda angosta $n(t)$.
6. La densidad espectral cruzada de las componentes en fase y en cuadratura del ruido de banda angosta $n(t)$ es puramente imaginaria, de acuerdo con

$$\begin{aligned} S_{N_I N_Q}(f) &= -S_{N_Q N_I}(f) \\ &= \begin{cases} j[S_N(f + f_c) - S_N(f - f_c)], & -B \le f \le B \\ 0, & \text{en otro caso} \end{cases} \end{aligned} \tag{1.102}$$

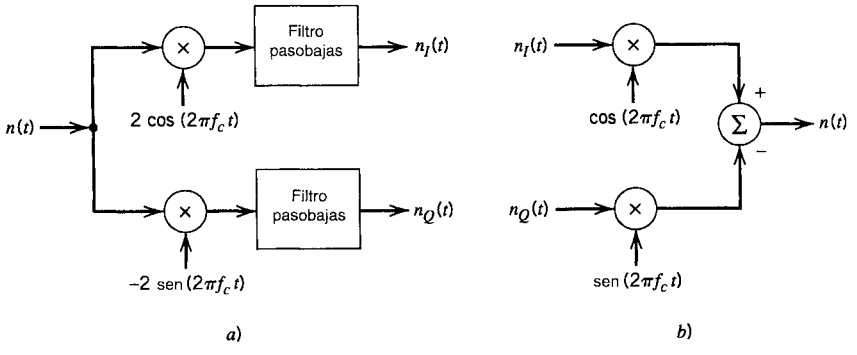


FIGURA 1.19 *a*) Extracción de las componentes en fase y en cuadratura de un proceso de banda angosta. *b*) Generación de un proceso de banda angosta a partir de sus componentes en fase y de cuadratura.

7. Si el ruido de banda angosta $n(t)$ es gaussiano y su densidad espectral de potencia $S_N(t)$ es simétrica en torno a la frecuencia de banda media f_c, entonces la componente en fase $n_I(t)$ y la componente en cuadratura $n_Q(t)$ son estadísticamente independientes.

Para un análisis adicional de estas propiedades, se remite al lector a los problemas 1.28 y 1.29.

▶ EJEMPLO 1.12 Ruido blanco con filtrado pasobanda ideal

Considere un ruido gaussiano blanco de media cero y densidad espectral de potencia $N_0/2$, que pasa a través de un filtro pasobanda ideal de respuesta en magnitud pasobanda igual a uno, frecuencia de banda media f_c y ancho de banda $2B$. La característica de la densidad espectral de potencia del ruido filtrado $n(t)$ será, en consecuencia, como se muestra en la figura 1.20*a*. El problema es determinar las funciones de autocorrelación $n(t)$ y sus componentes en fase y en cuadratura.

La función de autocorrelación de $n(t)$ es la transformada inversa de Fourier de la característica de densidad espectral de potencia que se ilustra en la figura 1.20*a*:

$$\begin{aligned} R_N(\tau) &= \int_{-f_c-B}^{-f_c+B} \frac{N_0}{2} \exp(j2\pi f\tau)\, df + \int_{f_c-B}^{f_c+B} \frac{N_0}{2} \exp(j2\pi f\tau)\, df \\ &= N_0 B \operatorname{senc}(2B\tau)[\exp(-j2\pi f_c\tau) + \exp(j2\pi f_c\tau)] \\ &= 2N_0 B \operatorname{senc}(2B\tau)\cos(2\pi f_c\tau) \end{aligned} \tag{1.103}$$

la cual se grafica en la figura 1.20*b*.

La característica de la densidad espectral de la figura 1.20*a* es simétrica alrededor de $\pm f_c$. Por tanto, encontramos que la característica de la densidad espectral correspondiente de la compo-

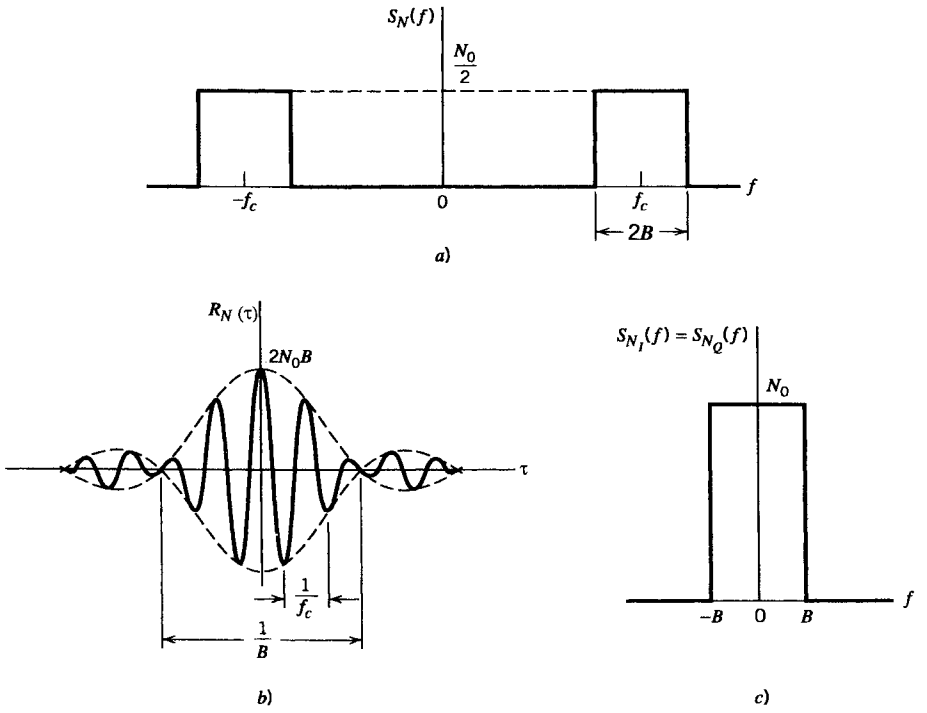


FIGURA 1.20 Características de ruido blanco con filtrado pasobanda ideal. *a*) Densidad espectral de potencia. *b*) Función de autocorrelación. *c*) Densidad espectral de potencia de las componentes en fase y en cuadratura.

nente de ruido en fase $n_I(t)$ o de la componente en cuadratura $n_Q(t)$ es como se muestra en la figura 1.21c. Por tanto, la función de autocorrelación de $n_I(t)$ o $n_Q(t)$ es (véase el ejemplo 1.10):

$$R_{N_I}(\tau) = R_{N_Q}(\tau) = 2N_0 B \ \text{senc}(2B\tau) \tag{1.104}$$

◀

1.12 *Representación del ruido de banda angosta en términos de las componentes de la envolvente y de la fase*

En la sección 1.11 consideramos la representación de un ruido de banda angosta $n(t)$ en términos de sus componentes en fase y en cuadratura. Es posible también representar el ruido $n(t)$ en términos de sus componentes de envolvente y fase de la manera siguiente:

$$n(t) = r(t)\cos[2\pi f_c t + \psi(t)] \tag{1.105}$$

donde

$$r(t) = \left[n_I^2(t) + n_Q^2(t)\right]^{1/2} \tag{1.106}$$

y

$$\psi(t) = \tan^{-1}\left[\frac{n_Q(t)}{n_I(t)}\right] \tag{1.107}$$

La función $r(t)$ se denomina la *envolvente* de $n(t)$, y la función $\psi(t)$ se denomina la *fase* de $n(t)$.

La envolvente $r(t)$ y la fase $\psi(t)$ son ambas funciones de muestreo de procesos aleatorios pasobajas. Como se ilustró en la figura 1.18*b*, el intervalo de tiempo entre dos picos sucesivos de la envolvente $r(t)$ es alrededor de $1/B$, donde $2B$ es el ancho de banda del ruido de banda angosta $n(t)$.

Las distribuciones de probabilidad de $r(t)$ y $\psi(t)$ pueden obtenerse de las correspondientes $n_I(t)$ y $n_Q(t)$ del modo siguiente. Dejemos que N_I y N_Q denoten las variables aleatorias obtenidas al observar (en algún tiempo fijo) el proceso aleatorio representado por las funciones de muestreo $n_I(t)$ y $n_Q(t)$, respectivamente. Advertimos que N_I y N_Q son variables aleatorias gaussianas independientes de media cero y varianza σ^2, por lo que es posible expresar su función de densidad de probabilidad conjunta mediante

$$f_{N_I,N_Q}(n_I,n_Q) = \frac{2}{2\pi\sigma^2}\exp\left(-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right) \tag{1.108}$$

En consecuencia, la probabilidad del evento conjunto de que N_I se encuentre entre n_I y $n_I + dn_I$ y de que N_Q se encuentre entre n_Q y $n_Q + dn_Q$ (es decir, el par de variables aleatorias N_I y N_Q se encuentre conjuntamente dentro del área sombreada de la figura 1.21*a*) está dada por

$$f_{N_I,N_Q}(n_I,n_Q)\,dn_I\,dn_Q = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right)dn_I\,dn_Q \tag{1.109}$$

Defina la transformación (vea la figura 1.21*a*)

$$n_I = r\cos\psi \tag{1.110}$$
$$n_Q = r\,\text{sen}\,\psi \tag{1.111}$$

En un sentido restrictivo, podemos igualar las dos áreas incrementales que se presentan sombreadas en las figuras 1.21*a* y 1.21*b* y escribir consecuentemente

$$dn_I\,dn_Q = r\,dr\,d\psi \tag{1.112}$$

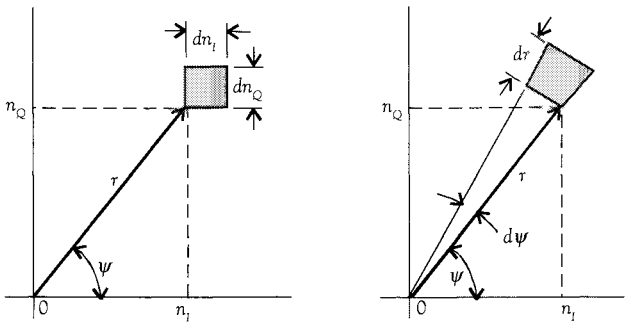


FIGURA 1.21 Ilustración del sistema de coordenadas para la representación del ruido de banda angosta: *a*) En términos de las componentes en fase y en cuadratura. *b*) En términos de la envolvente y la fase.

Después de esto, dejemos que R y Ψ denoten las variables aleatorias que se obtuvieron al observar (en algún tiempo t) el proceso aleatorio representado por la envolvente $r(t)$ y la fase $\psi(t)$, respectivamente. En ese caso, al sustituir las ecuaciones (1.110)-(1.112) en la (1.109), encontramos que la probabilidad de las variables aleatorias R y Ψ que se encuentran conjuntamente dentro del área sombreada de la figura 1.21*b* es igual a

$$\frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)dr\,d\psi$$

Esto es, la función de densidad de probabilidad conjunta de R y Ψ es

$$f_{R,\Psi}(r,\psi)=\frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{1.113}$$

Esta función de densidad de probabilidad es independiente del ángulo ψ, lo que significa que las variables aleatorias R y Ψ son estadísticamente independientes. En consecuencia, es posible expresar $f_{R,\Psi}(r,\psi)$ como el producto de $f_R(r)$ y $f_\Psi(\psi)$. En particular, la variable aleatoria Ψ que representa la fase está *distribuida uniformemente* dentro del intervalo de 0 a 2π, como se indica mediante

$$f_\Psi(\psi)=\begin{cases}\dfrac{1}{2\pi}, & 0\le\psi\le 2\pi\\ 0, & \text{en otro caso}\end{cases} \tag{1.114}$$

Esto deja la función de densidad de probabilidad de la variable aleatoria R como

$$f_R(r)=\begin{cases}\dfrac{r}{\sigma^2}\exp\left(-\dfrac{r^2}{2\sigma^2}\right), & r\ge 0\\ 0, & \text{en otro caso}\end{cases} \tag{1.115}$$

donde σ^2 es la varianza del ruido de banda angosta original $n(t)$. Una variable aleatoria que tiene la función de densidad de probabilidad de la ecuación (1.115) se dice que seguirá una *distribución de Rayleigh*.[8]

Por conveniencia de la representación gráfica, sea

$$v=\frac{r}{\sigma} \tag{1.116}$$

$$f_V(v)=\sigma f_R(r) \tag{1.117}$$

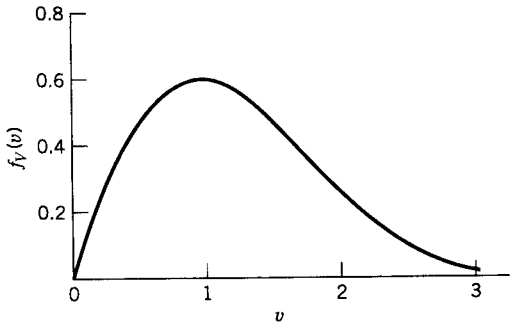


FIGURA 1.22 Distribución de Rayleigh normalizada.

En ese caso es posible reescribir la distribución de la ecuación de Rayleigh de la ecuación (1.115) en la forma *normalizada*

$$f_V(v) = \begin{cases} v \exp\left(-\frac{v^2}{2}\right), & v \geq 0 \\ 0, & \text{en otra parte} \end{cases} \tag{1.118}$$

La ecuación (1.118) se grafica en la figura 1.22. El valor pico de la distribución $f_V(v)$ ocurre en $v = 1$ y es igual a 0.607. Advierta también que, a diferencia de la distribución gaussiana, la de Rayleigh es cero para valores negativos de v. Ésta es la razón por la que la envolvente $r(t)$ puede tomar únicamente valores no negativos.

1.13 *Onda seno más ruido de banda angosta*

Suponga a continuación que agregamos la onda senoidal $A\cos(2\pi f_c t)$ al ruido de banda angosta $n(t)$, donde A y f_c son constantes. Suponemos que la frecuencia de la onda senoidal es la misma que la frecuencia de la portadora nominal del ruido. Una función de muestreo de la onda senoidal más el ruido se expresa entonces mediante

$$x(t) = A\cos(2\pi f_c t) + n(t) \tag{1.119}$$

Representando el ruido de banda angosta $n(t)$ en términos de sus componentes en fase y en cuadratura, es posible escribir

$$x(t) = n_I'(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \tag{1.120}$$

donde

$$n_I'(t) = A + n_I(t) \tag{1.121}$$

Suponemos que $n(t)$ es gaussiano como media cero y varianza σ^2. En consecuencia, podemos establecer lo siguiente:

1. Tanto $n_I'(t)$ como $n_Q(t)$ son gaussianas y estadísticamente independientes.
2. La media de $n_I'(t)$ es A y la de $n_Q(t)$ es cero.
3. La varianza tanto de $n_I'(t)$ como de $n_Q(t)$ es σ^2.

Por tanto, es factible expresar la función de densidad de probabilidad conjunta de las variables aleatorias N_I' y N_Q, correspondientes a $n_I'(t)$ y $n_Q(t)$, del modo siguiente:

$$f_{N_I',N_Q}(n_I', n_Q) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(n_I' - A)^2 + n_Q^2}{2\sigma^2}\right] \tag{1.122}$$

Dejemos que $r(t)$ denote la envolvente de $x(t)$ y $\psi(t)$ denote su fase. De acuerdo con la ecuación (1.120), encontramos que

$$r(t) = \left\{[n_I'(t)]^2 + n_Q^2(t)\right\}^{1/2} \tag{1.123}$$

y

$$\psi(t) = \tan^{-1}\left[\frac{n_Q(t)}{n_I'(t)}\right] \tag{1.124}$$

Siguiendo un procedimiento similar al que se describió en la sección 1.12 para la deducción de la distribución de Rayleigh, descubrimos que la función de densidad de probabilidad conjunta de las variables aleatorias R y Ψ, correspondientes a $r(t)$ y $\psi(t)$ para algún tiempo fijo t, está dada por

$$f_{R,\Psi}(r, \psi) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2 + A^2 - 2Ar\cos\psi}{2\sigma^2}\right) \tag{1.125}$$

Sin embargo, vemos que en este caso no es posible expresar la función de densidad de probabilidad conjunta $f_{R,\Psi}(r, \psi)$ como un producto de $f_R(r)f_\Psi(\psi)$. Esto se debe a que ahora tenemos un término que incluye los valores de ambas variables aleatorias multiplicados juntos como $r\cos\psi$. Por tanto, R y Ψ son variables aleatorias dependientes para valores distintos de cero de la amplitud A de la componente de la onda senoidal.

Nos interesa, en particular, la función de densidad de R. Para determinarla, integramos la ecuación (1.125) sobre todos los valores posibles de ψ obteniendo la densidad marginal

$$\begin{aligned} f_R(r) &= \int_0^{2\pi} f_{R,\Psi}(r, \psi)\, d\psi \\ &= \frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2 + A^2}{2\sigma^2}\right)\int_0^{2\pi} \exp\left(\frac{Ar}{\sigma^2}\cos\psi\right) d\psi \end{aligned} \tag{1.126}$$

La integral en el segundo miembro de la ecuación (1.126) puede identificarse en términos de la integral de definición para la *función de Bessel modificada del primer tipo de orden cero* (vea el apéndice 3); esto es,

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(x\cos\psi)\, d\psi \tag{1.127}$$

De tal modo, dejando $x = Ar/\sigma^2$, es posible reescribir la ecuación (1.126) en la forma compacta:

$$f_R(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{Ar}{\sigma^2}\right) \tag{1.128}$$

Esta relación se denomina la *distribución de Rician.*[9]

Como con la distribución de Rayleigh, la representación gráfica de la distribución de Rician se simplifica poniendo

$$v = \frac{r}{\sigma} \tag{1.129}$$

$$a = \frac{A}{\sigma} \tag{1.130}$$

$$f_V(v) = \sigma f_R(r) \tag{1.131}$$

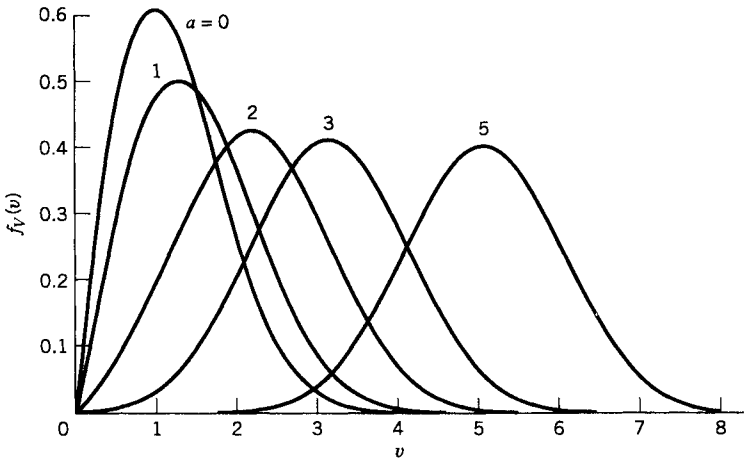


FIGURA 1.23 Distribución de Rician normalizada.

En ese caso, es posible expresar la distribución de Rician de la ecuación (1.128) en la forma *normalizada*

$$f_V(v) = v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av) \tag{1.132}$$

la cual se grafica en la figura 1.23 para los valores 0, 1, 2, 3, 5, del parámetro a. Con base en estas curvas, establecemos las siguientes observaciones:

1. Cuando a es cero, la distribución de Rician se reduce a la distribución de Rayleigh.
2. La distribución de la envolvente es aproximadamente gaussiana en la vecindad de $v = a$, cuando a es grande, esto es, cuando la amplitud A de la onda seno es grande en comparación con σ, la raíz cuadrada de la potencia del ruido $n(t)$.

1.14 *Experimentos de computadora: canal de desvanecimiento plano*

En esta sección utilizaremos simulaciones de computadora para estudiar un *canal de trayectoria múltiple* caracterizado por *distribución de Rayleigh,* de la cual se presentan ejemplos en las comunicaciones inalámbricas y en la transmisión de radio de largo alcance por medio de la ionosfera. El desvanecimiento ocurre debido a la interferencia entre versiones diferentes de la señal transmitida, las cuales llegan al receptor en tiempos consecuentemente diferentes. El resultado neto es que la señal recibida puede variar en forma extensa tanto en amplitud como en fase. Bajo ciertas condiciones, la naturaleza estadística variable en el tiempo de la envolvente de la señal recibida se describe bastante bien por medio de una distribución de Rayleigh, como se demuestra aquí.

La figura 1.24 representa un modelo de un canal de trayectoria múltiple. Éste consiste en una gran colección de *dispersores* posicionados al azar en el espacio, mediante los cuales un solo haz incidente se convierte en un número correspondientemente grande de haces dispersados en la antena receptora. La señal transmitida se iguala a $A\cos(2\pi f_c t)$. Se supone que todos los haces dispersados viajan a la misma velocidad media. Sin embargo, difieren entre sí en amplitud y fase por virtud de las diferencias en pérdidas y retardos de trayectoria. De tal modo, el haz dispersado k-ésimo está dado por $A_k \cos(2\pi f_c t + \Theta_k)$, donde la amplitud A_k y la fase Θ_k son variables aleatorias que cambian lentamente con el tiempo. Además, todas las Θ_k son independientes entre sí y se distribuyen de

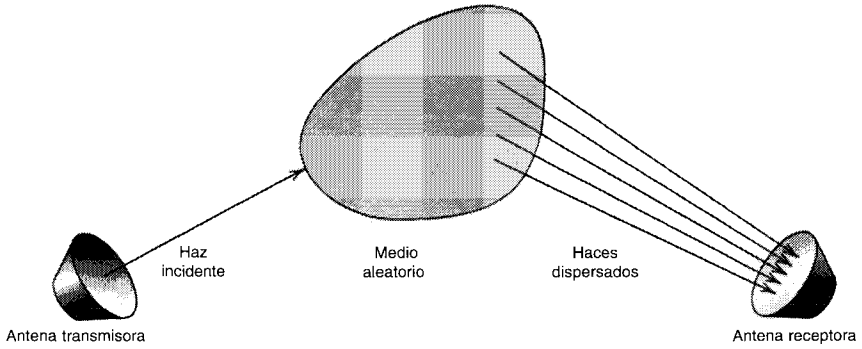


FIGURA 1.24 Modelo de un canal de trayectoria múltiple.

manera uniforme dentro del intervalo $[0, 2\pi]$. El tipo de desvanecimiento exhibido por el canal de trayectoria múltiple descrito aquí se conoce como "disminución plana" debido a que las características espectrales de la señal transmitida se reservan por completo a la salida del canal. Sin embargo, la intensidad de la salida del canal cambia con el tiempo a causa de las fluctuaciones aleatorias en la ganancia del canal provocadas por el fenómeno de trayectoria múltiple.

Sumando las contribuciones de todos los dispersores, supuestos en número N, es posible expresar el proceso aleatorio que presenta la señal recibida como

$$X(t) = \sum_{k=1}^{N} A_k \cos(2\pi f_c t + \Theta_k) \tag{1.133}$$

que puede reescribirse en la forma equivalente

$$X(t) = X_I \cos(2\pi f_c t) - X_Q \operatorname{sen}(2\pi f_c t) \tag{1.134}$$

donde X_I y X_Q se definen respectivamente por medio de

$$X_I = \sum_{k=1}^{N} A_k \cos\Theta_k \tag{1.135}$$

y

$$X_Q = \sum_{k=1}^{N} A_k \operatorname{sen}\Theta_k \tag{1.136}$$

Por conveniencia en la presentación y sin perder generalidad, resulta factible suponer que A_k se encuentra en el intervalo $[-1, 1]$ para toda k.

Experimento 1. Distribuciones gaussianas

A partir del teorema del límite central advertimos que cuando el número de dispersores, N, tiende a infinito, tanto X_I como X_Q deben tender a variables aleatorias gaussianas. Para probar la validez de este planteamiento, se calculan las distribuciones de probabilidad de la componente en fase X_I y de la componente en cuadratura X_Q para $N = 10$, 100, 1000 y 10,000. La prueba de la validez del teorema de límite central necesita una medida de la calidad del ajuste para comprobar la equivalencia de la distribución de probabilidad medida de los datos muestreados para N variable con la

distribución gaussiana teórica. Una forma de llevar a cabo esta prueba es utilizar momentos centrales de una distribución (hasta el orden 4) para definir los siguientes dos parámetros:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \tag{1.137}$$

y

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{1.138}$$

donde μ_2, μ_3 y μ_4 son el segundo, tercero y cuarto momentos centrales, respectivamente. Los parámetros β_1 y β_2 proporcionan en conjunto una medida de la *asimetría* de la distribución bajo prueba. Cuanto más cercanos son los valores de β_1 y β_2 para la distribución medida a los correspondientes para la distribución teórica, tanto mejor es el ajuste. Para una variable aleatoria gaussiana X de media μ_X y varianza σ_X^2 tenemos

$$\mu_2 = \sigma_X^2$$

$$\mu_3 = 0$$

$$\mu_4 = 3\sigma_X^4$$

lo cual produce

$$\beta_1 = 0$$

y

$$\beta_2 = 3$$

La tabla 1.1 presenta los valores de β_1 y β_2 calculados tanto para la componente en fase X_I como para la componente en cuadratura X_Q para N variable. Al comparar estos valores con los correspondientes a la distribución gaussiana, vemos en forma clara que cuando aumenta el número de dispersores N, las distribuciones tanto de X_I como de X_Q tienden a la distribución gaussiana de media cero en concordancia con el teorema de límite central.

TABLA 1.1 ***Valores β para las componentes en fase y en cuadratura***

a) Distribución medida

		Número de dispersores, N			
		10	100	1000	10,000
Componente en fase X_I	β_1	0.2443	0.0255	0.0065	0.0003
	β_2	2.1567	2.8759	2.8587	3.0075
Componente de cuadratura X_Q	β_1	0.0874	0.0017	0.0004	0.0000
	β_2	1.9621	2.7109	3.1663	3.0135

b) Distribución teórica: gaussiana

$$\beta_1 = 0$$

$$\beta_2 = 3$$

Experimento 2. Distribución de Rayleigh

En la ecuación (1.134), el proceso aleatorio $X(t)$ se expresa en términos de sus componentes en fase y en cuadratura. De manera equivalente, es posible expresar $X(t)$ en términos de su envolvente y fase como

$$X(t) = R\cos(2\pi f_c t + \Psi) \tag{1.139}$$

donde

$$R = \sqrt{X_I^2 + X_Q^2} \tag{1.140}$$

y

$$\Psi = \tan^{-1}\left(\frac{X_Q}{X_I}\right) \tag{1.141}$$

Advierta que en los experimentos considerados aquí la componente en fase X_I, la componente en cuadratura X_Q, la envolvente R y la fase Ψ son todas independientes del tiempo.

Si X_I y X_Q se aproximan a variables aleatorias gaussianas para N creciente, entonces a partir de la teoría que se presentó en la sección 1.12 notamos que la envolvente R se aproximará a una distribución de Rayleigh y que la fase Ψ tenderá a una distribución uniforme. En la figura 1.25 presentamos la función de densidad de probabilidad real de r para datos generados en el caso de $N = 10{,}000$, con el cálculo de 100 histogramas y 100 promedios totales. Esta figura comprende también la curva teórica. Existe una estrecha concordancia entre estas dos curvas, lo que sostiene la afirmación de que la envolvente R de la señal recibida tiende a una distribución de Rayleigh.

La figura 1.26 ilustra el efecto del desvanecimiento de Rayleigh en la forma de onda de la señal recibida $x(t)$, una función de muestreo de $X(t)$, en el caso de una señal transmitida senoidal con amplitud unitaria (es decir, $A = 1$) y frecuencia $f_c = 1$ MHz. Específicamente, la señal transmitida y la señal recibida correspondiente se muestran en las partes *a* y *b* de la figura 1.26, respectivamente. Al comparar estas dos formas de onda, observamos que la transmisión a través

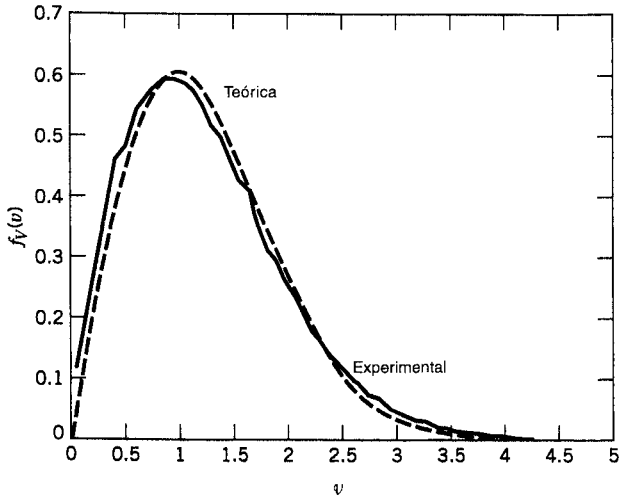


FIGURA 1.25 Función de densidad de probabilidad de la envolvente del proceso aleatorio $X(t)$: comparación de la teoría y el experimento.

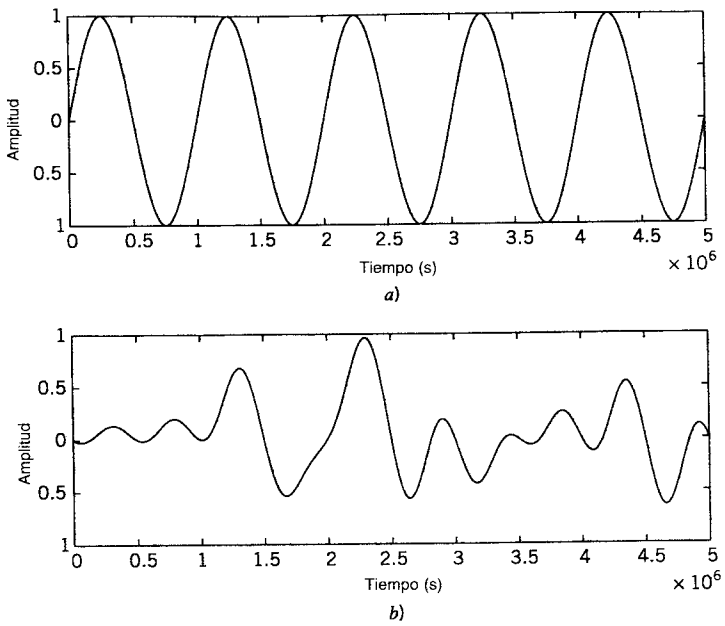


FIGURA 1.26 Efecto del desvanecimiento de Rayleigh en una onda senoidal. *a*) Onda senoidal de entrada. *b*) Forma de onda de la señal resultante.

del canal de trayectoria múltiple de la figura 1.24 ha producido una señal recibida cuyas componentes de amplitud y fase varían en forma aleatoria con el tiempo, como se esperaba.

1.15 *Resumen y discusión*

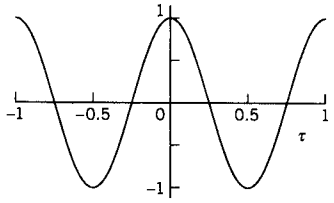
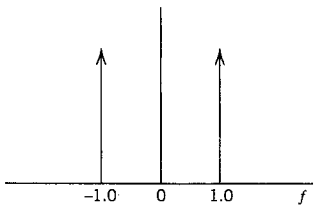
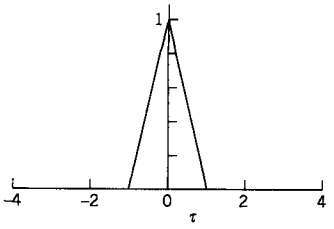
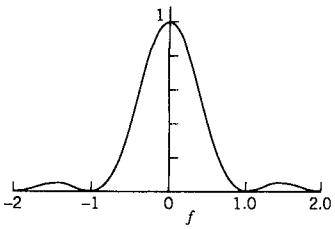
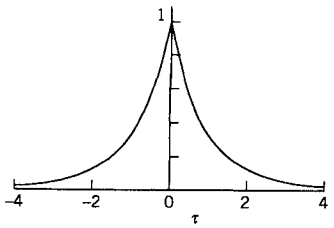
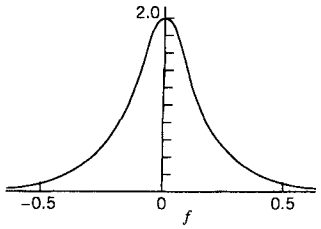
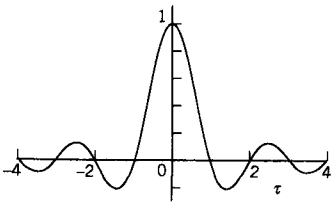
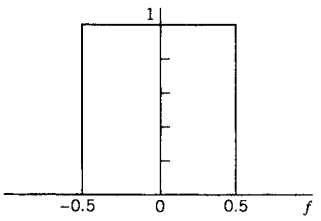
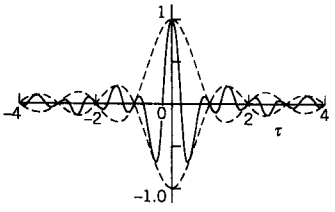
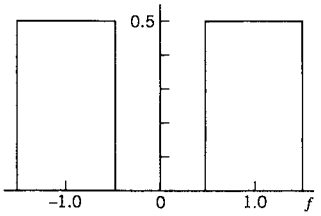
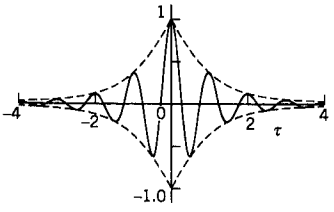
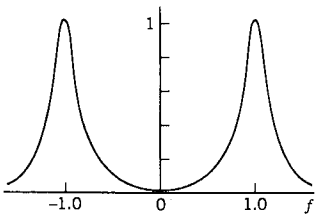
Gran parte del material presentado en este capítulo tiene que ver con la caracterización de una clase particular de procesos aleatorios conocidos como estacionarios y ergódicos. La implicación de la estacionariedad (en el sentido amplio) es que es posible que desarrollemos una descripción parcial de un proceso aleatorio en términos de dos parámetros promediados en el conjunto: 1) una media que es independiente del tiempo y 2) una función de autocorrelación que depende únicamente de la diferencia entre los tiempos en los que se realizan dos observaciones del proceso.[10] La ergodicidad nos permite utilizar promedios de tiempo como "estimaciones" de estos parámetros. Los promedios de tiempo se calculan utilizando una función de muestreo (es decir, una sola realización) del proceso aleatorio.

Otro parámetro importante de un proceso aleatorio es la densidad espectral de potencia. La función de autocorrelación y la densidad espectral de potencia constituyen un par de transformadas de Fourier. Las fórmulas que definen la densidad espectral de potencia en términos de la función de autocorrelación y viceversa, se conocen como las relaciones de Einstein-Wiener-Khintchine.

En la tabla 1.2 presentamos un resumen gráfico de las funciones de autocorrelación y de las densidades espectrales de potencia de procesos aleatorios importantes. Todos los procesos que se describen en esta tabla se suponen que tienen media cero y varianza unitaria. Esta tabla debe proporcionar al lector una idea relativa a 1) la interacción entre la función de autocorrelación y la densidad espectral de potencia de un proceso aleatorio, y 2) el papel del filtrado lineal en la forma de la función de autocorrelación o, de manera equivalente, la densidad espectral de potencia de un proceso de ruido blanco.

La última parte del capítulo abordó un proceso de ruido que es gaussiano y de banda angosta, el cual es del tipo de ruido filtrado que se encuentra en el extremo frontal de una forma idealizada de receptor de comunicación. Calidad de gaussiano significa que la variable aleatoria obtenida al obser-

TABLA 1.2 ***Resumen gráfico de las funciones de autocorrelación y de las densidades espectrales de potencia de procesos aleatorios de media cero y varianza unitaria***

Tipo de proceso, $X(t)$	*Función de autocorrelación, $R_X(\tau)$*	*Densidad espectral de potencia, $S_X(f)$*
Proceso senoidal de frecuencia unitaria y fase aleatoria		
Onda binaria aleatoria de duración de símbolo unitaria		
Ruido blanco con filtrado pasobajas RC		
Ruido blanco con filtrado pasobajas ideal		
Ruido blanco con filtrado pasobanda ideal		
Ruido blanco con filtrado RLC		

var la salida del filtro en algún tiempo fijo tiene una distribución gaussiana. La naturaleza de banda angosta del ruido quiere decir que éste quizá se represente en términos de una componente en fase y de una en cuadratura. Las dos componentes son procesos gaussianos pasobajas, cada una con media cero y varianza igual a la del ruido de banda angosta original. En forma alternativa, es posible representar un ruido de banda angosta gaussiano en términos de una envolvente de distribución de Rayleigh y una fase distribuida uniformemente. Cada una de estas representaciones tiene su propia área de aplicación específica, como se muestra en los capítulos subsecuentes del libro.

NOTAS Y BIBLIOGRAFÍA

1. Para un tratamiento riguroso de los procesos aleatorios, vea los libros clásicos de Doob [38], Loève [95] y Cramér y Leadbetter [34].

2. Existe otra clase importante de procesos aleatorios encontrados comúnmente en la práctica, de los cuales la media y la función de autocorrelación exhiben *periodicidad,* como en

$$\mu_X(t_1 + T) = \mu_X(t_1)$$
$$R_X(t_1 + T, t_2 + T) = R_X(t_1, t_2)$$

para toda t_1 y t_2. Un proceso aleatorio $X(t)$ que satisface este par de condiciones, se dice que es *cicloestacionario* (en el sentido amplio). La modelación del proceso $X(t)$ como cicloestacionario añade una nueva dimensión, a saber, el periodo T a la descripción parcial del proceso. Ejemplos de procesos cicloestacionarios incluyen una señal de televisión obtenida mediante la exploración óptica de trama de un campo de video aleatorio y un proceso modulado obtenido al variar la amplitud, la fase o la frecuencia de una portadora senoidal. Para una explicación detallada de los procesos cicloestacionarios, consulte Franks [43], pp. 204-214, y el artículo de Gardner y Franks [258].

3. Tradicionalmente, las ecuaciones (1.42) y (1.43) se han conocido en la literatura técnica como las relaciones de Wiener-Khintchine en reconocimiento del trabajo pionero realizado por Norbert Wiener y A.I. Khintchine; en lo que respecta a los artículos originales, vea Wiener, y Khintchine [295]. El descubrimiento de un artículo olvidado de Albert Einstein acerca del análisis de las series de tiempo (entregado a una reunión de febrero de 1914 de la Sociedad Física Suiza en Basilea) revela que Einstein había analizado la función de autocorrelación y su relación con el contenido espectral de las series del tiempo muchos años antes que Wiener y Khintchine. Una traducción al inglés del artículo de Einstein se reproduce en el *IEEE ASSP Magazine,* vol. 4. Este número particular también contiene artículos de W.A. Gardner y A.M. Yaglom, los cuales se elaboraron con base en el trabajo original de Einstein.

4. Para mayores detalles de la estimación espectral de potencia, vea Blachman y Tukey [21], Box y Jenkins [25], Marple y Kay.

5. La distribución gaussiana y el proceso gaussiano asociado reciben ese nombre en honor al gran matemático C.F. Gauss. A la edad de 18 años, Gauss inventó el método de *mínimos cuadrados* para determinar el mejor valor de una secuencia de mediciones de alguna cantidad. Gauss utilizó después el método de mínimos cuadrados para ajustar órbitas de planetas a las mediciones de datos, un procedimiento que se publicó en 1809 en su libro titulado *Teoría del Movimiento de Cuerpos Celestes.* En conexión con el error de observación, desarrolló la *distribución gaussiana.* Ésta también se conoce como la *distribución normal.* En parte por razones históricas, los matemáticos emplean comúnmente el término normal, si bien los ingenieros y los físicos utilizan la mayor parte de las veces el término gaussiano.

6. En cuanto a un tratamiento detallado del ruido eléctrico, vea Van der Ziel [157] y la colección de artículos editados por Gupta [59].

   Un tratamiento introductorio del ruido de disparo se presenta en Helstrom [67]. Para un tratamiento más amplio, consulte el artículo de Yue, Luganani y Rice [409].

El ruido térmico lo estudió experimentalmente por primera vez J.B. Johnson en 1928, y por esta razón algunas veces se denomina *ruido Johnson*. Los experimentos de Johnson los confirmó teóricamente Nyquist [332].

7. La calidad de ruidoso de un receptor también puede medirse en términos de la llamada *figura de ruido*. La relación entre la figura de ruido y la temperatura de ruido equivalente se desarrolla en el capítulo 8.
8. La distribución de Rayleigh recibe este nombre en honor al físico inglés J.W. Strutt, Lord Rayleigh.
9. La distribución de Rician recibe este nombre en honor a Stephen O. Rice, por la aportación original de la cual informó en un par de artículos publicados en 1944 y 1945, que se reproducen en Wax [164].
10. La caracterización estadística de los sistemas de comunicación presentada en este libro se confina a los primeros dos momentos, la media y función de autocorrelación (equivalente a función de autocovarianza) del proceso aleatorio pertinente. Sin embargo, cuando un proceso aleatorio se transmite a través de un sistema no lineal, se encuentra contenida información valiosa en momentos de orden superior del proceso de salida resultante. Los parámetros utilizados para caracterizar los momentos de orden superior en el dominio del tiempo se denominan *cumulantes*, y sus transformadas de Fourier multidominantes se denominan *poliespectros*. En cuanto a una explicación de los cumulantes de orden superior y los poliespectros y su estimación, vea el artículo de Nikias y Raghuveer.

PROBLEMAS

Estacionariedad y ergodicidad

1.1 Considere un proceso aleatorio $X(t)$ definido por medio de

$$X(t) = \text{sen}(2\pi f_c t)$$

en el que la frecuencia f_c es una variable aleatoria definida uniformemente sobre el intervalo $[0, W]$. Demuestre que $X(t)$ es no estacionario. *Sugerencia*: examine funciones de muestreo específicas del proceso aleatorio $X(t)$ para la frecuencia $f = W/4, W/2$, y W, entre otras posibilidades.

1.2 Considere el proceso senoidal

$$X(t) = A\cos(2\pi f_c t)$$

donde la frecuencia f_c es constante y la amplitud A está distribuida uniformemente:

$$f_A(a) = \begin{cases} 1, & 0 \le a \le 1 \\ 0, & \text{en otro caso} \end{cases}$$

Determine si el proceso es o no estrictamente estacionario.

1.3 Un proceso aleatorio $X(t)$ está definido por

$$X(t) = A\cos(2\pi f_c t)$$

donde A es una variable aleatoria distribuida en forma gaussiana de media cero y varianza σ_A^2. Este proceso aleatorio se aplica a un integrador ideal y se produce la salida

$$Y(t) = \int_0^t X(\tau)\, d\tau$$

a) Determine la función de densidad de probabilidad de la salida $Y(t)$ en un tiempo particular t_k.
b) Determine si $Y(t)$ es o no estacionaria.
c) Determine si $Y(t)$ es o no ergódica.

1.4 Sean X y Y variables aleatorias distribuidas en forma gaussiana estadísticamente independientes, cada una con media cero y varianza unitaria. Defina el proceso gaussiano

$$Z(t) = X\cos(2\pi t) + Y\,\text{sen}(2\pi t)$$

a) Determine la función de densidad de probabilidad conjunta de las variables aleatorias $Z(t_1)$ y $Z(t_2)$ obtenida al observar $Z(t)$ en los tiempos t_1 y t_2, respectivamente.

b) ¿El proceso $Z(t)$ es estacionario? ¿Por qué?

Funciones de correlación y de densidad espectral

1.5 Compruebe las siguientes dos propiedades de la función de autocorrelación $R_X(\tau)$ de un proceso aleatorio $X(t)$:

a) Si $X(t)$ contiene una componente de cd igual a A, entonces $R_X(\tau)$ contendrá una componente constante igual a A^2.

b) Si $X(t)$ contiene una componente senoidal, entonces $R_X(\tau)$ contendrá también una componente senoidal $x(t)$ de la misma frecuencia.

1.6 La onda cuadrada $x(t)$ de la figura P1.6 de amplitud constante A, periodo T_0 y retraso t_d, representa la función de muestreo de un proceso aleatorio $X(t)$. El retraso es aleatorio y se describe por medio de la función de densidad de probabilidad

$$f_{T_d}(t_d)=\begin{cases}\dfrac{1}{T_0}, & -\dfrac{1}{2}T_0 \le t_d \le \dfrac{1}{2}T_0\\ 0, & \text{en otro caso}\end{cases}$$

a) Determine la función de densidad de probabilidad aleatoria $X(t_k)$ obtenida al observar el proceso aleatorio $X(t)$ en el tiempo t_k.

b) Determine la media y la función de autocorrelación de $X(t)$ utilizando promediación total.

c) Determine la media y la función de autocorrelación de $X(t)$ utilizando promediación en el tiempo.

d) Establezca si $X(t)$ es o no estacionaria. ¿En qué sentido esto es ergódico?

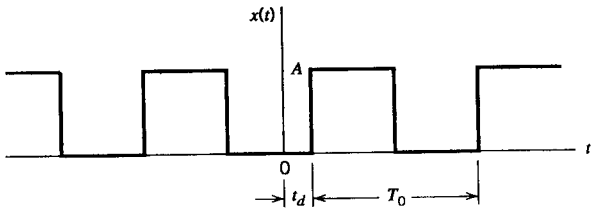


FIGURA P1.6

1.7 Una onda binaria está compuesta por una frecuencia aleatoria de símbolos 1 y 0, similar a la que se describe en el ejemplo 1.3, con una diferencia básica: el símbolo 1 se representa en este caso por medio de un pulso de amplitud A en volts y el símbolo 0 mediante cero volts. La totalidad de los demás parámetros son los mismos de antes. Demuestre que para esta nueva onda binaria aleatoria $X(t)$:

a) La función de autocorrelación es

$$R_X(\tau)=\begin{cases}\dfrac{A^2}{4}+\dfrac{A^2}{4}\left(1-\dfrac{|\tau|}{T}\right), & |\tau|<T\\ \dfrac{A^2}{4}, & |\tau|\ge T\end{cases}$$

b) La densidad espectral de potencia corresponde a

$$S_X(f)=\frac{A^2}{4}\delta(f)+\frac{A^2T}{4}\,\text{senc}^2(fT)$$

¿Cuál es el porcentaje de potencia contenida en la componente de cd en la onda binaria?

1.8 Un proceso aleatorio $Y(t)$ está compuesto por una componente de cd de $\sqrt{3/2}$ volts, una componente periódica $g(t)$ y una componente aleatoria $X(t)$. La función de autocorrelación $Y(t)$ se presenta en la figura P1.8.

a) ¿Cuál es la potencia promedio de la componente periódica $g(t)$?

b) ¿Cuál es la potencia promedio de la componente aleatoria $X(t)$?

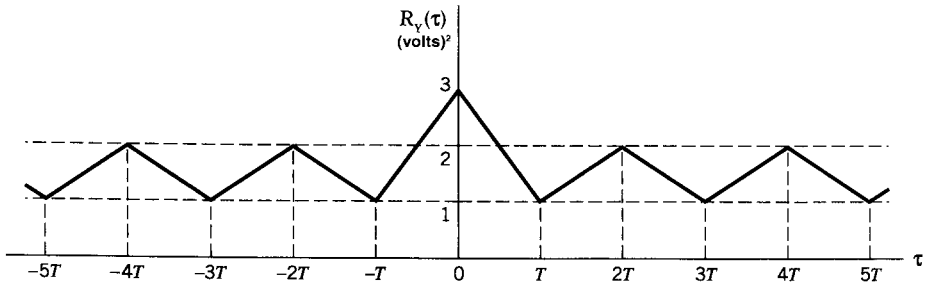


FIGURA P1.8

1.9 Considere un par de procesos estacionarios $X(t)$ y $Y(t)$. Demuestre que las correlaciones cruzadas $R_{XY}(\tau)$ y $R_{YX}(\tau)$ de estos procesos tienen las siguientes propiedades:

a) $R_{XY}(\tau) = R_{YX}(-\tau)$

b) $|R_{XY}(\tau)| \leq \frac{1}{2}[R_X(0) + R_Y(0)]$

donde $R_X(\tau)$ y $R_Y(\tau)$ son las funciones de autocorrelación de $X(t)$ y $Y(t)$, respectivamente.

1.10 Considere dos filtros lineales conectados en cascada como en la figura P1.10. Sea $X(t)$ un proceso estacionario con función de autocorrelación $R_X(\tau)$. El proceso aleatorio que aparece en la primera salida del filtro es $V(t)$ y el correspondiente a la salida del segundo filtro es $Y(t)$.

a) Encuentre la función de autocorrelación de $Y(t)$.

b) Determine la función de correlación cruzada $R_{VY}(\tau)$ de $V(t)$ y $Y(t)$.

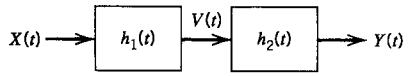


FIGURA P1.10

1.11 Se aplica un proceso estacionario $X(t)$ a un filtro lineal invariante con el tiempo de respuesta al impulso $h(t)$, produciéndose una salida $Y(t)$.

a) Demuestre que la función de correlación cruzada $R_{YX}(\tau)$ de la salida $Y(t)$ y de la entrada $X(t)$ es igual a la respuesta al impulso $h(\tau)$ convolucionada con la función de autocorrelación $R_X(\tau)$ de la entrada, tal como indica

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} h(u) R_X(\tau - u)\, du$$

Demuestre que la segunda función de correlación cruzada $R_{XY}(\tau)$ es igual a

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(-u) R_X(\tau - u)\, du$$

b) Determine las densidades espectrales cruzadas $S_{YX}(f)$ y $S_{XY}(f)$.

c) Suponiendo que $X(t)$ es un proceso de ruido blanco con media cero y densidad espectral de potencia $N_0/2$, demuestre que

$$R_{YX}(\tau) = \frac{N_0}{2} h(\tau)$$

Comente acerca del significado práctico de este resultado.

1.12 La densidad espectral de potencia de un proceso aleatorio $X(t)$ se presenta en la figura P1.12. Consta de una función delta de $f = 0$ y de una componente triangular.

a) Determine y dibuje la función de autocorrelación $R_X(\tau)$ de $X(t)$.

b) ¿Cuál es la potencia de cd contenida en $X(t)$?

c) ¿Cuál es la potencia de ca contenida en $X(t)$?

d) ¿Qué velocidades de muestreo producirán muestras no correlacionadas de $X(t)$? ¿Las muestras son estadísticamente independientes?

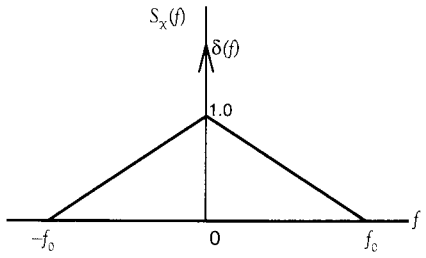


FIGURA P1.12

1.13 Un par de procesos de ruido $n_1(t)$ y $n_2(t)$ se relacionan por medio de

$$n_2(t) = n_1(t)\cos(2\pi f_c t + \theta) - n_1(t)\,\text{sen}(2\pi f_c t + \theta)$$

donde f_c es una constante y θ es el valor de la variable aleatoria Θ cuya función de densidad de probabilidades está definida por

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & 0 \le \theta \le 2\pi \\ 0, & \text{en otro caso} \end{cases}$$

El proceso de ruido $n_1(t)$ es estacionario y su densidad espectral de potencia es como se indica en la figura P1.13. Determine y grafique la densidad de potencia correspondiente de $n_2(t)$.

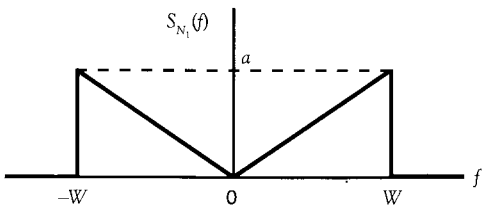


FIGURA P1.13

1.14 Una *señal telegráfica aleatoria* $X(t)$, caracterizada por la función de autocorrelación

$$R_X(\tau) = \exp(-2v|\tau|)$$

donde v es una constante, se aplica al filtro RC pasobajas de la figura P1.14. Determine la densidad espectral de potencia y la función de autocorrelación del proceso aleatorio a la salida del filtro.

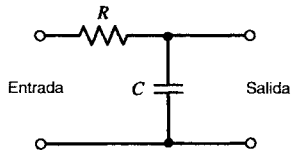


FIGURA P1.14

1.15 Un *integrador continuo* se define por

$$y(t) = \int_{t-T}^{t} x(\tau)\, d\tau$$

donde $x(t)$ es la entrada, $y(t)$ es la salida y T es el periodo de integración. Tanto $x(t)$ como $y(t)$ son las funciones de muestreo de los procesos estacionarios $X(t)$ y $Y(t)$, respectivamente. Demuestre que la densidad espectral de potencia de la salida del integrador se relaciona con la correspondiente a la entrada del integrador de acuerdo con

$$S_Y(f) = T^2 \operatorname{senc}^2(fT) S_X(f)$$

1.16 Un proceso estacionario $X(t)$ de media cero se aplica a un filtro lineal cuya respuesta al impulso está definida por medio de una exponencial truncada:

$$h(t) = \begin{cases} ae^{-at}, & 0 \le t \le T \\ 0, & \text{en otro caso} \end{cases}$$

Demuestre que la densidad espectral de potencia de la salida del filtro $Y(t)$ está definida por

$$S_Y(f) = \frac{a^2}{a^2 + 4\pi^2 f^2}(1 - 2\exp(-aT)\cos(2\pi fT) + \exp(-2aT))S_X(f)$$

donde $S_X(f)$ es la densidad espectral de potencia de la entrada del filtro.

1.17 La salida de un oscilador se describe por medio de

$$X(t) = A\cos(2\pi ft - \Theta)$$

donde A es una constante y f y Θ son variables independientes. La función de densidad de probabilidad de Θ está definida por

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & 0 \le \theta \le 2\pi \\ 0, & \text{en otro caso} \end{cases}$$

Encuentre la densidad espectral de potencia de $X(t)$ en términos de la función de densidad de probabilidad de la frecuencia f. ¿Qué sucede con esta densidad espectral de potencia cuando la frecuencia f toma un valor constante?

Proceso gaussiano

1.18 Un proceso gaussiano estacionario $X(t)$ tiene media cero y densidad espectral de potencia $S_X(f)$. Determine la función de densidad de probabilidad de una variable aleatoria obtenida al observar el proceso $X(t)$ en algún tiempo t_k.

1.19 Un proceso gaussiano $X(t)$ de media cero y varianza σ_X^2 pasa a través de un rectificador de onda completa, el cual se describe mediante la relación de entrada-salida de la figura P1.19. Compruebe que la función de densidad de probabilidad de la variable aleatoria $Y(t_k)$, obtenida al observar el proceso aleatorio $Y(t)$ en la salida del rectificador en el tiempo t_k, es como sigue:

$$f_{Y(t_k)}(y) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_X} \exp\left(-\frac{y^2}{2\sigma_X^2}\right), & y \geq 0 \\ 0, & y < 0 \end{cases}$$

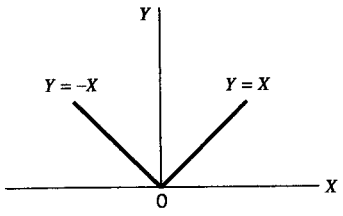


FIGURA P1.19

1.20 Sea $X(t)$ un proceso gaussiano estacionario de media cero con función de autocorrelación $R_X(\tau)$. Este proceso se aplica a un dispositivo de ley cuadrática, el cual se define mediante la relación de entrada-salida

$$Y(t) = X^2(t)$$

donde $Y(t)$ es la salida.

a) Demuestre que la media de $Y(t)$ es $R_X(0)$.

b) Demuestre que la función de autocovarianza de $Y(t)$ es $2R_X^2(\tau)$.

1.21 Un proceso estacionario gaussiano $X(t)$ con media μ_X y varianza σ_X^2 pasa a través de dos filtros lineales con respuestas al impulso $h_1(t)$ y $h_2(t)$, produciendo los procesos $Y(t)$ y $Z(t)$, como se ilustra en la figura P1.21.

a) Determine la función de densidad de probabilidad conjunta de las variables aleatorias $Y(t_1)$ y $Z(t_2)$.

b) ¿Qué condiciones son necesarias y suficientes para asegurar que $Y(t_1)$ y $Z(t_2)$ son estadísticamente independientes?

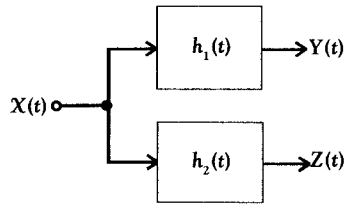


FIGURA P1.21

1.22 Un proceso gaussiano estacionario $X(t)$ con media cero y densidad espectral de potencia $S_X(f)$ se aplica a un filtro lineal cuya respuesta al impulso $h(t)$ se presenta en la figura P1.22. Se toma una muestra Y del proceso aleatorio en la salida del filtro en el tiempo T.

a) Determine la media y la varianza de Y.

b) ¿Cuál es la función de densidad de probabilidad de Y?

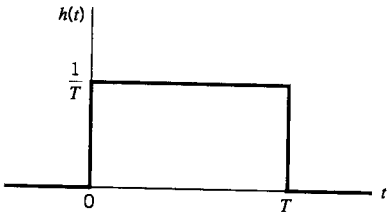


FIGURA P1.22

Ruido

1.23 Considere el proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$ que se aplica al filtro RL pasoaltas que se muestra en la figura P1.23.

a) Determine la función de autocorrelación y la densidad espectral de potencia del proceso aleatorio en la salida del filtro.

b) ¿Cuáles son la media y la varianza de esta salida?

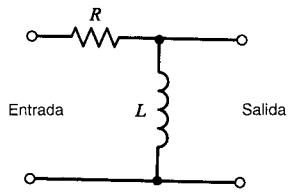


FIGURA P1.23

1.24 Un ruido blanco $w(t)$ de densidad espectral de potencia $N_0/2$ se aplica a un filtro *Butterworth* pasobajas de orden n, cuya respuesta en magnitud está definida por medio de

$$|H(f)| = \frac{1}{\left[1+(f/f_0)^{2n}\right]^{1/2}}$$

a) Determine el ancho de banda equivalente de ruido para este filtro pasobajas. (Vea el apéndice 2 para la definición del ancho de banda equivalente de ruido.)

b) ¿Cuál es el valor límite del ancho de banda equivalente de ruido cuando n tiende a infinito?

1.25 El proceso de ruido de disparo $X(t)$ definido por la ecuación (1.86) es estacionario. ¿Por qué?

1.26 Se aplica ruido gaussiano blanco de media cero y densidad espectral de potencia $N_0/2$ al esquema de filtrado que se indica en la figura 1.26*a*. Las respuestas en frecuencia de estos dos filtros se muestran en la figura 1.26*b*. El ruido en la salida del filtro pasobajas se denota mediante $n(t)$.

a) Encuentre la densidad espectral de potencia y la función de autocorrelación de $n(t)$.

b) Determine la media y la varianza de $n(t)$.

c) ¿Cuál es la velocidad a la que $n(t)$ puede muestrearse de manera que las muestras que se produzcan esencialmente no estén correlacionadas?

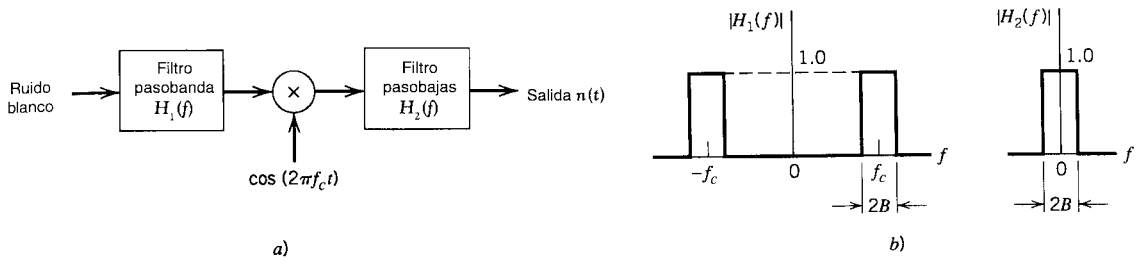


FIGURA P1.26

1.27 Considere que $X(t)$ es un proceso estacionario con media cero, función de autocorrelación $R_X(\tau)$ y densidad espectral de potencia $S_X(f)$. Se nos pide determinar un filtro lineal con respuesta al impulso $h(t)$, de manera que la salida del filtro tenga las mismas características estadísticas que $X(t)$ cuando la entrada es ruido blanco de densidad espectral de potencia $N_0/2$.

a) Determine la condición que la respuesta al impulso $h(t)$ debe satisfacer para cumplir este requerimiento.

b) ¿Cuál es la condición correspondiente con respecto a la respuesta en frecuencia $H(f)$ del filtro?

Ruido de banda angosta

1.28 En el analizador de ruido de la figura 1.19*a*, los filtros pasobajas son ideales con un ancho de banda igual a la mitad del ruido de banda angosta $n(t)$ aplicado a la entrada. Mediante este esquema, derive los siguientes resultados:

a) La ecuación (1.101), que define las densidades espectrales de potencia de la componente de ruido en fase $n_I(t)$ y la componente de ruido de cuadratura $n_Q(t)$ en términos de la densidad espectral de potencia $n(t)$.

b) La ecuación (1.102), que define las densidades espectrales cruzadas de $n_I(t)$ y $n_Q(t)$.

1.29 Suponga que el ruido de banda angosta $n(t)$ es gaussiano y que su densidad espectral de potencia $S_N(f)$ es simétrica en torno a la frecuencia de banda media f_c. Demuestre que las componentes en fase y en cuadratura de $n(t)$ son estadísticamente independientes.

1.30 La densidad espectral de potencia de un ruido de banda angosta $n(t)$ es como se indica en la figura P1.30. La frecuencia de portadora corresponde a 5 Hz.

a) Determine las densidades espectrales de potencia de las componentes en fase y en cuadratura de $n(t)$.

b) Encuentre sus densidades espectrales cruzadas.

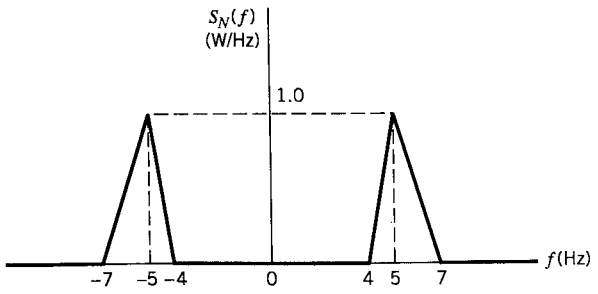


FIGURA P1.30

1.31 Considere un ruido gaussiano $n(t)$ con media cero y la densidad espectral de potencia $S_N(f)$ que se muestra en la figura P1.31.

a) Encuentre la función de densidad de probabilidad de la envolvente $n(t)$.

b) ¿Cuáles son la media y la varianza de esta envolvente?

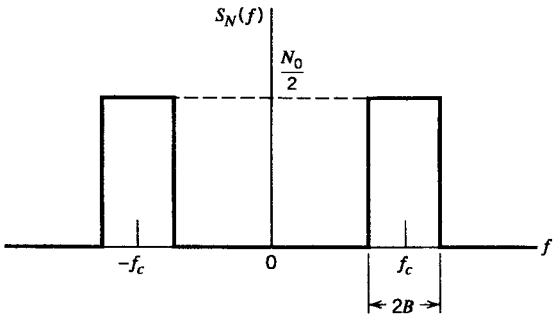


FIGURA P1.31

Experimentos por computadora

1.32 En este experimento por computadora estudiamos la caracterización estadística de un proceso aleatorio $X(t)$ definido por

$$X(t) = A\cos(2\pi f_c t + \Theta) + W(t)$$

donde la fase Θ de la componente senoidal es una variable aleatoria distribuida de manera uniforme sobre el intervalo $[-\pi, \pi]$, y $W(t)$ es una componente de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$. Las dos componentes de $X(t)$ son estadísticamente independientes; en consecuencia, la función de autocorrelación de $X(t)$ es

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau) + \frac{N_0}{2}\delta(\tau)$$

Esta ecuación muestra que para $|\tau| > 0$ la función de autocorrelación $R_X(\tau)$ tiene la misma forma de onda senoidal que la componente de la señal de $X(t)$.

El propósito de este experimento de computadora es efectuar el cálculo de $R_X(t)$ utilizando dos métodos diferentes:

a) *Promediación de totales.* Se generan M = 50 realizaciones escogidas al azar del proceso $X(t)$. De ese modo, el producto $x(t + \tau)x(t)$ para algún tiempo fijo t, donde $x(t)$ es una realización de $X(t)$. Se repite el cálculo de $x(t + \tau)x(t)$ para las M realizaciones de $X(t)$, y de ahí en adelante se calcula el promedio de estos cálculos sobre M. Se repite esta secuencia de cálculos para diferentes valores de τ.

b) *Promediación en el tiempo.* Se calcula la función de correlación promediada en el tiempo

$$R_x(\tau, T) = \frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)\,dt$$

donde $x(t)$ es una realización particular de $X(t)$, y $2T$ es el intervalo de observación total. Para este cálculo, use el par de transformadas de Fourier:

$$R_x(\tau, T) \rightleftharpoons \frac{1}{2T}|X_T(f)|^2$$

donde $|X_T(f)|^2/2T$ es el periodograma del proceso $X(t)$. Específicamente, calcule la transformada de Fourier $X_T(f)$ de la función con ventana en el tiempo

$$x_T(t) = \begin{cases} x(t), & -T \le t \le T \\ 0, & \text{en otro caso} \end{cases}$$

A partir de aquí calcule la transformada inversa de Fourier de $|X_T(f)|^2/2T$.

Compare los resultados de su cálculo de $R_X(\tau)$ utilizando estos dos procedimientos.

1.33 En este experimento de computadora continuamos el estudio del canal de trayectoria múltiple descrito en la sección 1.14. De modo específico, considere la situación en la que la señal recibida incluya una componente de línea de vista, como se indica mediante

$$X(t) = \sum_{k=1}^{N} A_k \cos(2\pi f_c t + \Theta_k) + a\cos(2\pi f_c t)$$

donde $a\cos(2\pi f_c t)$ es la componente recibida directamente. Siguiendo el material que se presenta en la sección 1.14, calcule la envolvente de $X(t)$ para $N = 10{,}000$, y $a = 0, 1, 2, 3, 5$. Compare sus resultados con la distribución de Rician estudiada en la sección 1.13.

CAPÍTULO 2

MODULACIÓN DE ONDA CONTINUA

En este capítulo estudiamos la modulación de onda continua, la cual resulta básica para la operación de los sistemas de comunicación analógicos. El capítulo se divide en dos partes relacionadas. En la primera estudiamos las descripciones en el dominio del tiempo y en el dominio de la frecuencia de dos familias básicas de modulación de onda continua:

- *Modulación de amplitud; en ésta, la amplitud de una portadora senoidal se varía de acuerdo con una señal del mensaje entrante.*
- *Modulación angular, en la cual la frecuencia o fase instantánea de la portadora senoidal se varía de acuerdo con la señal de mensaje.*

En la segunda parte del capítulo se analizan los efectos del ruido del canal en el desempeño de los receptores relativos a estos esquemas de modulación.

Se subrayan las ventajas y desventajas de los diferentes métodos de modulación de onda continua a la luz del material que se presenta en este capítulo.

2.1 *Introducción*

El objetivo de un sistema de comunicación es transmitir *señales que contienen información* a través de un canal de comunicación que separa al transmisor del receptor. Las señales que llevan información también se conocen como *señales de banda base*. El término *banda base* se utiliza para designar la banda de frecuencias que representa la señal original como la entrega de la fuente de información. El uso apropiado del canal de comunicación requiere un corrimiento del intervalo de las frecuencias de banda base en otros intervalos de frecuencias adecuados para la transmisión, así como en un correspondiente corrimiento de regreso al intervalo de frecuencia original después de la recepción. Por ejemplo, un sistema de radio debe operar con frecuencias de 30 kHz y superiores, en tanto que la señal de la banda base suele contener frecuencias en la gama de las frecuencias de audio, por lo que alguna forma de corrimiento de la banda de frecuencia debe emplearse para que el sistema opere satisfactoriamente. Un corrimiento del intervalo de frecuencias en una señal se logra utilizando *modulación*, la cual se define como *el proceso mediante el que alguna característica de la portadora se varía de acuerdo con una onda (señal) moduladora*. Una forma común de portadora es una *onda senoidal*, en cuyo caso hablamos de un proceso de *modulación de onda continua*.[1] La señal de banda base se conoce como *onda moduladora* y el resultado del proceso de modulación se denomina *onda modulada*. La modulación se realiza en el extremo de transmisión del sistema de comunicación. En el extremo receptor del sistema, generalmente requerimos que se recupere la señal de banda base original, lo cual se consigue al utilizar un proceso conocido como *demodulación*, que es el inverso del proceso de modulación.

En términos del procesamiento básico de señales, encontramos que el transmisor de un sistema de comunicación analógico consta de un modulador y que el receptor está compuesto por

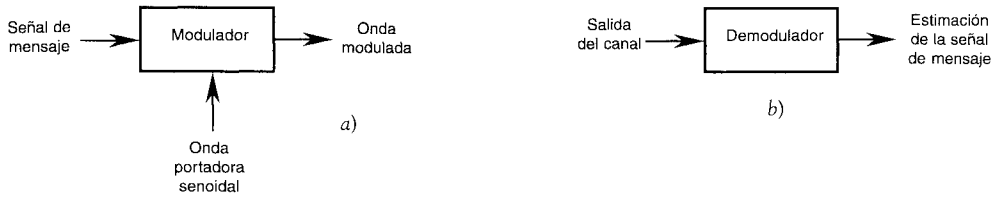


FIGURA 2.1 Componentes de un sistema de modulación de onda continua: *a*) transmisor y *b*) receptor.

un demodulador, en la forma que se describe en la figura 2.1. Además de la señal recibida del transmisor, la entrada del receptor incluye ruido del canal. La degradación en el desempeño del receptor por causa del ruido del canal se determina mediante el tipo de modulación que se utiliza.

En este capítulo estudiamos dos familias de sistemas de modulación de onda continua, a saber: *modulación de amplitud* y *modulación angular*. En el primer tipo se varía la amplitud de la onda portadora senoidal de acuerdo con la señal de la banda base. En la modulación angular, el ángulo de la onda portadora senoidal se varía en concordancia con la señal de la banda base. La figura 2.2 exhibe las formas de onda de las señales modulada de amplitud y modulada angular para el caso de la modulación senoidal. Las partes *a* y *b* de la figura muestran la portadora senoidal y las ondas moduladoras, respectivamente. Las partes *c* y *d* presentan las correspondientes ondas modulada

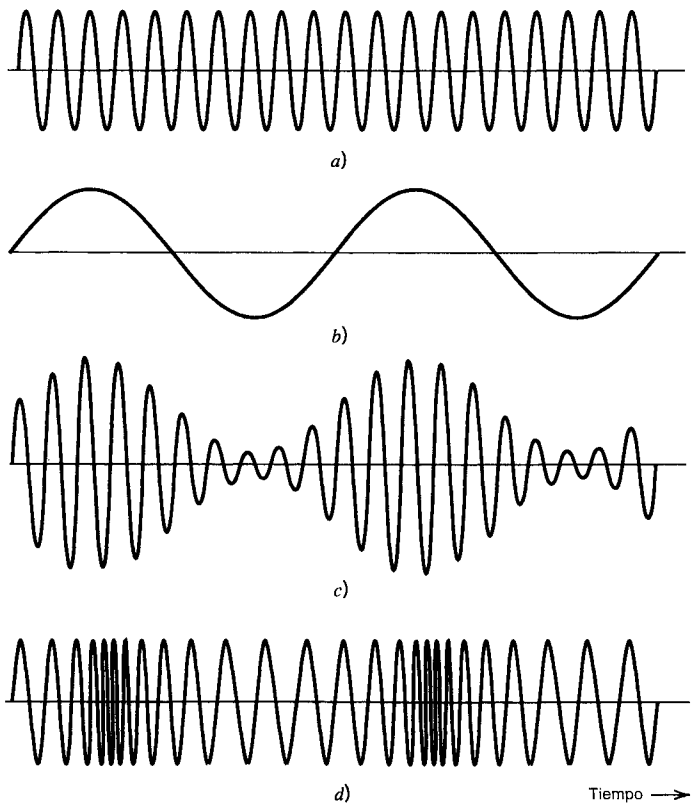


FIGURA 2.2 Ilustración de señales de AM y FM producidas por un tono. *a*) Onda portadora. *b*) Señal moduladora senoidal. *c*) Señal modulada en amplitud. *d*) Señal modulada en frecuencia.

de amplitud y modulada de frecuencia, respectivamente; la modulación de frecuencia es una modulación angular. Esta figura ilustra con claridad las diferencias básicas entre la modulación de amplitud y la modulación angular, las cuales se explican en lo que sigue.

2.2 *Modulación de amplitud*

Considere una *onda portadora senoidal* $c(t)$ definida por

$$c(t) = A_c \cos(2\pi f_c t) \tag{2.1}$$

donde A_c es la *amplitud de la portadora* y f_c es la *frecuencia de la portadora*. Para simplificar la exposición sin afectar los resultados que se obtengan y las conclusiones que se alcancen, hemos supuesto que la fase de la onda portadora es cero en la ecuación (2.1). Dejemos que $m(t)$ denote la señal de la banda base que lleva la especificación del mensaje. La fuente de la onda portadora $c(t)$ es físicamente independiente de la fuente responsable de generar $m(t)$. La *modulación de amplitud (AM) se define como el proceso en el que la amplitud de la onda portadora $c(t)$ se varía en torno a un valor medio, linealmente con la señal de la banda base $m(t)$*. Consecuentemente, una onda de amplitud modulada (AM) puede describirse, en su forma más general, como una función del tiempo de la manera siguiente:

$$s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t) \tag{2.2}$$

donde k_a es una constante denominada la *sensibilidad de amplitud* del modulador responsable de la generación de la señal modulada $s(t)$. Por lo común, la amplitud A_c de la portadora y la señal del mensaje $m(t)$ se miden en volts, en cuyo caso k_a se mide en volt^{-1}.

La figura 2.3*a* muestra una señal de banda base $m(t)$ y las figuras 2.3*b* y 2.3*c* presentan la correspondiente onda de AM, $s(t)$ para los dos valores de sensibilidad de amplitud k_a y la amplitud de portadora $A_c = 1$ volt. Observamos que la *envolvente* de $s(t)$ tiene esencialmente la misma forma que la señal de banda base $m(t)$ siempre que se cumplan dos requerimientos:

1. La amplitud de $k_a m(t)$ siempre es menor que la unidad, esto es,

$$|k_a m(t)| < 1 \quad \text{para todo } t \tag{2.3}$$

Esta condición se ilustra en la figura 2.3*b*; ésta asegura que la función $1 + k_a m(t)$ sea siempre positiva y, puesto que la envolvente es una función positiva, podemos expresar la envolvente de la onda de AM, $s(t)$ de la ecuación (2.2) como $A_c[1 + k_a m(t)]$. Cuando la sensibilidad de amplitud k_a del modulador es suficientemente grande para hacer que $|k_a m(t)| > 1$ para todo t, la onda portadora se *sobremodula*, lo que resulta en inversiones de fase de la portadora siempre que el factor $1 + k_a m(t)$ cruza por cero. La onda modulada exhibe en ese caso *distorsión de envolvente*, como en la figura 2.3*c*. Por tanto, es patente que al evitar la sobremodulación, se mantiene una relación uno a uno entre la envolvente de la onda AM y la onda moduladora para todos los valores de tiempo, una característica útil, como veremos más adelante. El valor máximo absoluto de $k_a m(t)$ multiplicado por 100 se conoce como el *porcentaje de modulación*.

2. La frecuencia de la portadora f_c es mucho mayor que la componente de frecuencia más alta W de la señal del mensaje $m(t)$, esto es,

$$f_c \gg W \tag{2.4}$$

Llamamos a W el *ancho de banda del mensaje*. Si la condición de la ecuación (2.4) no se satisface, no es posible visualizar (y consecuentemente detectar) una envolvente en forma satisfactoria.

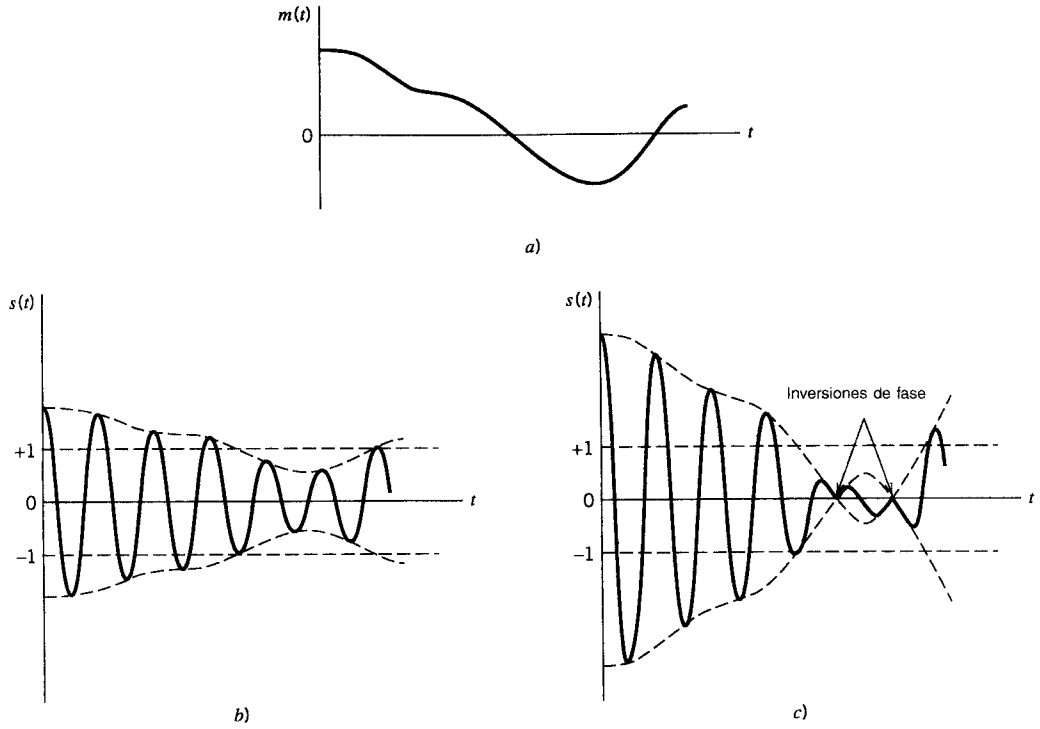


FIGURA 2.3 Ilustración del proceso de modulación de amplitud. *a*) Señal de banda base $m(t)$. *b*) Onda de AM para $|k_a m(t)| < 1$ para todo t. *c*) Onda de AM para $|k_a m(t)| > 1$ para algún t.

De acuerdo con la ecuación (2.2), encontramos que la transformada de Fourier de la onda AM $s(t)$ está dada por

$$S(f) = \frac{A_c}{2}[\delta(f - f_c) + \delta(f + f_c)] + \frac{k_a A_c}{2}[M(f - f_c) + M(f + f_c)] \tag{2.5}$$

Suponga que la señal de banda base $m(t)$ está limitada en banda en el intervalo $-W \leq f \leq W$, como en la figura 2.4*a*. La forma del espectro que se muestra en esta figura atiende sólo a fines ilustrativos. Encontramos a partir de la ecuación (2.5) que el espectro $S(f)$ de la onda de AM es como se muestra en la figura 2.4*b* para el caso en el que $f_c > W$. Este espectro consta de dos funciones delta dadas por el factor $A_c/2$ que se presentan en $\pm f_c$, y de dos versiones del espectro de la banda base trasladado en la frecuencia por $\pm f_c$ y escalado en amplitud por $k_a A_c/2$. A partir del espectro de la figura 2.4*b*, advertimos lo siguiente:

1. Como resultado del proceso de modulación, el espectro de la señal del mensaje $m(t)$ para frecuencias negativas que se extiende desde $-W$ hasta 0 se vuelve por completo visible para frecuencias positivas (es decir, medible), siempre que la frecuencia de la portadora cumpla la condición $f_c > W$; aquí se encuentra la importancia de la idea de frecuencias "negativas".

2. Para frecuencias positivas, la porción del espectro de la onda de AM que se ubica sobre la frecuencia de la portadora f_c se conoce como la *banda lateral superior*, en tanto que la porción simétrica debajo de f_c se denomina la *banda lateral inferior*. En el caso de frecuencias negativas, la banda lateral superior se representa por medio de la porción del espectro por debajo de $-f_c$ y la banda lateral inferior por la parte de arriba de $-f_c$. La condición $f_c > W$ asegura que no se traslapen las bandas laterales.

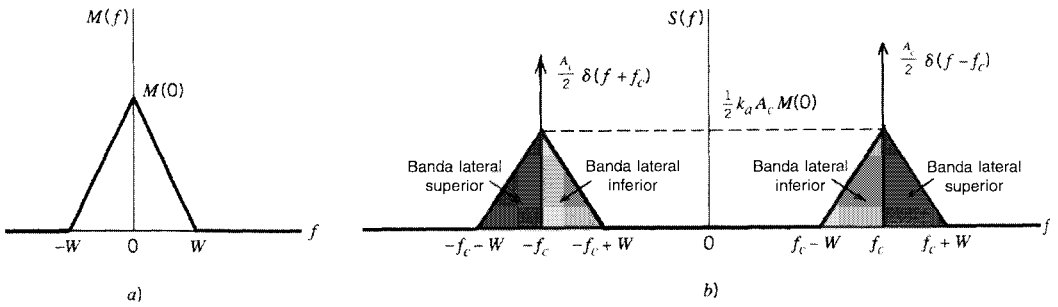


FIGURA 2.4 *a*) Espectro de la señal de banda base. *b*) Espectro de la onda de AM.

3. Para frecuencias positivas, la componente de frecuencia más alta de la onda de AM es igual a $f_c + W$, y la componente de frecuencia más baja corresponde a $f_c - W$. La diferencia entre estas dos frecuencias define el *ancho de banda de transmisión* B_T para una onda de AM, que es exactamente el doble del ancho de banda del mensaje W, esto es,

$$B_T = 2W \tag{2.6}$$

VIRTUDES Y LIMITACIONES DE LA MODULACIÓN DE AMPLITUD

La modulación de amplitud es el método más antiguo con el que se realiza la modulación. Su mayor virtud corresponde a la simplicidad de su puesta en práctica:

- En el transmisor, la modulación de amplitud se lleva a cabo utilizando un dispositivo no lineal. Por ejemplo, en el *modulador de conmutación* que se discute en el problema 2.3, la suma combinada de la señal del mensaje y la onda portadora se aplica a un diodo, con la amplitud de la portadora lo suficientemente grande para oscilar a lo largo de la curva característica del diodo. El análisis de Fourier del voltaje que se crea en los extremos de una carga resistiva revela la generación de una componente de AM, que es posible extraer por medio de un filtro pasobanda.
- En el receptor, la demodulación de amplitud se realiza también utilizando un dispositivo no lineal. Por ejemplo, es posible utilizar un circuito simple y altamente efectivo conocido como el *detector de envolvente*, que se explica en el problema 2.5. El circuito consta de un diodo conectado en serie con la combinación en paralelo de un capacitor y de un resistor de carga. Cierta versión de este circuito se encuentra en los receptores de radio de AM más comerciales. Siempre que la frecuencia de la portadora sea lo suficientemente alta y que el porcentaje de modulación sea menor que 100%, la salida del demodulador que se genera en los extremos del resistor de carga es casi la misma que la envolvente de la onda de AM entrante, de ahí el nombre de "detector de envolvente".

Sin embargo, recuerde que la potencia transmitida y el ancho de banda del canal son nuestros dos recursos de comunicación principales y que éstos deben utilizarse de manera eficiente. En este contexto encontramos que la forma estándar de la modulación de amplitud definida en la ecuación (2.2) sufre de dos limitaciones fundamentales:

1. La *modulación de amplitud desperdicia potencia*. La onda portadora $c(t)$ es por completo independiente de la señal que contiene información $m(t)$. Por tanto, la transmisión de la onda portadora representa un desperdicio de potencia, lo cual significa que en la modulación de amplitud sólo una fracción de la potencia transmitida total es en realidad afectada por $m(t)$.

2. La *modulación de amplitud desperdicia ancho de banda.* Las bandas laterales superior e inferior de una onda de AM se relacionan unívocamente entre sí por virtud de su simetría en torno a la frecuencia de la portadora; por tanto, dados los espectros de magnitud y fase de cualquier banda lateral, podemos determinar en forma única la otra. Esto quiere decir que en lo que se refiere a la transmisión de información, únicamente es necesaria una banda lateral, y en consecuencia, el canal de comunicación sólo necesita proporcionar el mismo ancho de banda que el de la señal de la banda base. A la luz de esta observación, la modulación de amplitud desperdicia ancho de banda, puesto que requiere un ancho de banda de transmisión igual al doble del relativo al mensaje.

Para superar estas limitaciones, debemos recurrir a ciertas modificaciones: suprimir la portadora y modificar las bandas laterales de la onda de AM. Estas modificaciones producen naturalmente una mayor complejidad del sistema. En efecto, cambiamos la complejidad del sistema por un mejor uso de los recursos de comunicación. La base de este compromiso es la modulación lineal, la cual se explica en la sección siguiente. En estricto sentido, la modulación de amplitud completa no califica como modulación lineal debido a la presencia de la onda portadora.

2.3 *Esquemas de modulación lineal*

En su forma más general, la modulación lineal se define por medio de

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\,\mathrm{sen}(2\pi f_c t) \tag{2.7}$$

donde $s_I(t)$ es la *componente en fase* de la onda modulada $s(t)$, y $s_Q(t)$ es su *componente de cuadratura*. La ecuación (2.7) se reconoce como la representación canónica de una señal de banda angosta, la cual se discute en detalle en el apéndice 2. En la modulación lineal, tanto $s_I(t)$ como $s_Q(t)$ son señales pasobajas que se relacionan linealmente con la señal del mensaje $m(t)$.

En realidad, dependiendo de cómo se definen estas dos componentes de $s(t)$, es posible identificar tres tipos de modulación lineal que incluyen una sola señal del mensaje:

1. *Modulación de doble banda lateral-portadora suprimida (DSB-SC)*, donde únicamente se transmiten las bandas laterales superior e inferior.
2. *Modulación de banda lateral única (SSB)*, donde únicamente se transmite una banda lateral (la banda lateral inferior o la superior).
3. *Modulación de banda lateral residual (VSB)*, donde sólo se transmite un residuo (es decir, una traza) de una de las bandas laterales y una versión modificada correspondiente de la otra banda lateral.

En la tabla 2.1 se presenta un resumen de las definiciones de estas tres formas especiales de modulación lineal. Hay dos puntos importantes que señalar de la tabla 2.1:

1. La componente en fase $s_I(t)$ depende exclusivamente de la señal de mensaje $m(t)$.
2. La componente de cuadratura $s_Q(t)$ es una versión *filtrada* de $m(t)$. Esta modificación espectral de la onda modulada $s(t)$ se debe únicamente a $s_Q(t)$.

Para ser más específicos, el papel de la componente de cuadratura (si está presente) es meramente el de interferir con la componente en fase, a fin de reducir o eliminar potencia en una de las bandas laterales de la señal modulada $s(t)$, dependiendo de cómo se defina la componente en cuadratura.

TABLA 2.1 *Formas diferentes de la modulación lineal*

Tipo de modulación	*Componente en fase* $s_I(t)$	*Componente en cuadratura* $s_Q(t)$	*Comentarios*
DSB-SC	$m(t)$	0	$m(t)$ = señal del mensaje
SSB:[a]			
a) Banda lateral superior transmitida	$\frac{1}{2}m(t)$	$\frac{1}{2}\hat{m}(t)$	$\hat{m}(t)$ = transformada de Hilbert de $m(t)$
b) Banda lateral inferior transmitida	$\frac{1}{2}m(t)$	$-\frac{1}{2}\hat{m}(t)$	
VSB:			$m'(t)$ = salida del filtro de respuesta en frecuencia $H_Q(f)$ debida a $m(t)$. Para la definición de $H_Q(f)$, vea la ecuación (2.16)
a) Banda lateral inferior residual transmitida	$\frac{1}{2}m(t)$	$\frac{1}{2}m'(t)$	
b) Banda lateral superior residual transmitida	$\frac{1}{2}m(t)$	$-\frac{1}{2}m'(t)$	

[a]Para la descripción matemática de una modulación de banda lateral única, vea el problema 2.16.

■ MODULACIÓN DE DOBLE BANDA LATERAL-PORTADORA SUPRIMIDA (DSB-SC)

Esta forma de modulación lineal se genera utilizando un *modulador de producto* que simplemente multiplica la señal del mensaje $m(t)$ por la onda portadora $A_c \cos(2\pi f_c t)$, como se ilustra en la figura 2.5*a*. De manera específica, escribimos

$$s(t) = A_c m(t) \cos(2\pi f_c t) \tag{2.8}$$

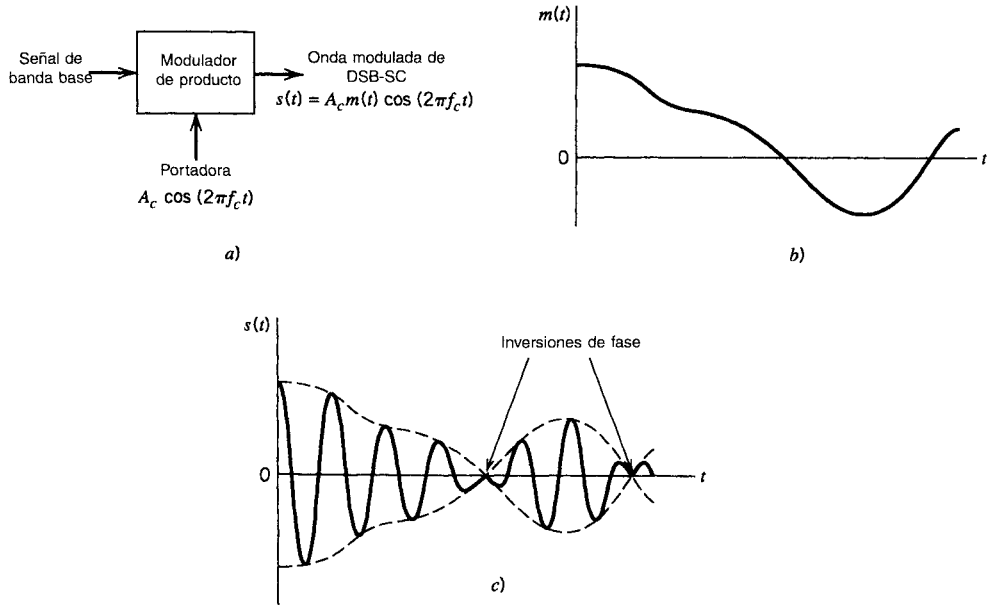


FIGURA 2.5 *a*) Diagrama de bloques del modulador de producto. *b*) Señal de banda base. *c*) Onda modulada de DSB-SC.

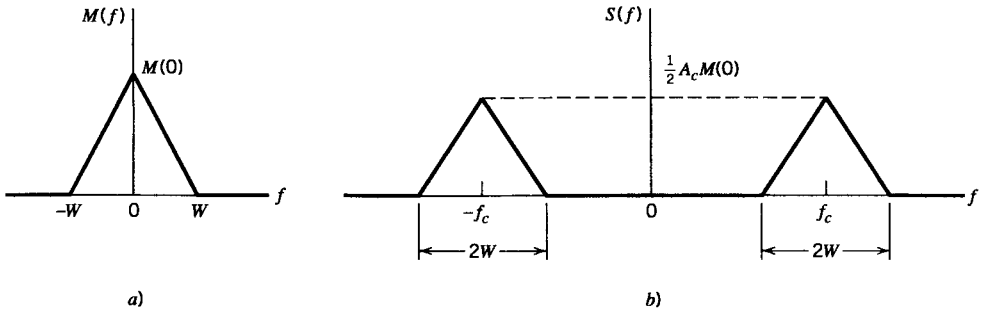


FIGURA 2.6 *a*) Espectro de la señal de banda base. *b*) Espectro de la onda modulada de DSB-SC.

La figura 2.5c presenta la señal modulada $s(t)$ para la forma de onda del mensaje arbitrario de la figura 2.5*b*. La señal modulada $s(t)$ experimenta una *inversión de fase* siempre que la señal del mensaje $m(t)$ pasa por cero. En consecuencia, la envolvente de una señal modulada DSB-SC es diferente de la señal del mensaje; lo anterior es distinto al caso de una onda de AM que tiene un porcentaje de modulación menor que 100%.

De acuerdo con la ecuación (2.8), la transformada de Fourier de $s(t)$ se obtiene como

$$S(f) = \frac{1}{2}A_c[M(f - f_c) + M(f + f_c)] \tag{2.9}$$

Para el caso en el que la señal de banda base $m(t)$ está limitada al intervalo $-W \leq f \leq W$, como en la figura 2.6*a*, encontramos consecuentemente que el espectro $S(f)$ de la onda de DSB-SC $s(t)$ es como se ilustra en la figura 2.6*b*. Salvo por el cambio en el factor de escala, el proceso de modulación simplemente *traslada* el espectro de la señal de banda base en $\pm f_c$. Desde luego, el ancho de banda de transmisión requerido por la modulación DSB-SC es el mismo que el correspondiente a la modulación de amplitud, a saber, $2W$.

■ DETECCIÓN COHERENTE

La señal de banda base $m(t)$ puede recuperarse en forma única a partir de la onda DSB-SC $s(t)$ multiplicando primero $s(t)$ con una onda senoidal generada localmente y haciendo pasar después el producto por un filtro pasobajas, como en la figura 2.7. Se supone que la señal del oscilador local es coherente o sincronizada, tanto en frecuencia como en fase, con la onda portadora $c(t)$ utilizada en el modulador de producto para generar $s(t)$. Este método de demodulación se conoce como *detección coherente* o *demodulación síncrona*.

Resulta instructivo deducir la detección coherente como un caso especial del proceso más general de demodulación al utilizar una señal de oscilador local de la misma frecuencia pero de diferencia de fase arbitraria ϕ, medida con respecto a la onda portadora $c(t)$. De esta manera, al

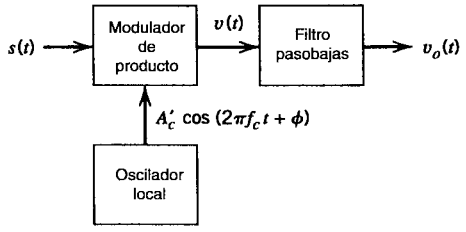


FIGURA 2.7 Detector coherente para demodular la onda modulada de DSB-SC.

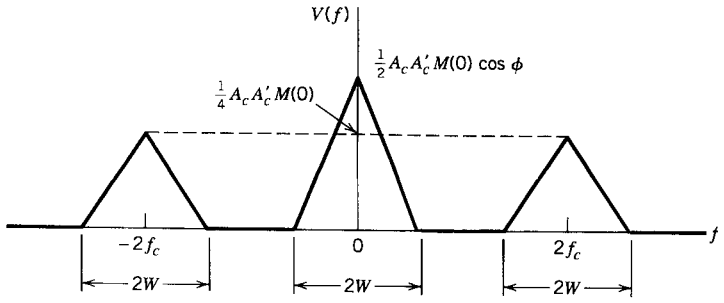


FIGURA 2.8 Ilustración del espectro de una salida de modulador de producto con una onda modulada de DSB-SC como entrada.

denotar la señal del oscilador local mediante $A_c' \cos(2\pi f_c t + \phi)$, y utilizar la ecuación (2.8) para la onda DSB-SC $s(t)$, encontramos que la salida del modulador de producto en la figura 2.7 es

$$\begin{aligned} v(t) &= A_c' \cos(2\pi f_c t + \phi) s(t) \\ &= A_c A_c' \cos(2\pi f_c t) \cos(2\pi f_c t + \phi) m(t) \\ &= \frac{1}{2} A_c A_c' \cos(4\pi f_c t + \phi) m(t) + \frac{1}{2} A_c A_c' \cos\phi \, m(t) \end{aligned} \tag{2.10}$$

El primer término en la ecuación (2.10) representa una señal modulada de DSB-SC con una frecuencia de portadora $2f_c$, en tanto que el segundo término es proporcional a la señal de banda base $m(t)$. Esto se ilustra adicionalmente por medio del espectro $V(f)$ que se presenta en la figura 2.8, donde se supone que la señal de banda base $m(t)$ está limitada al intervalo $-W \leq f \leq W$. En consecuencia, resulta claro que el primer término en la ecuación (2.10) se elimina mediante el filtro pasobajas de la figura 2.7, siempre que la frecuencia de corte de este filtro sea mayor que W pero menor que $2f_c - W$. Este requerimiento se satisface eligiendo $f_c > W$. A la salida del filtro obtenemos entonces una señal dada por

$$v_o(t) = \frac{1}{2} A_c A_c' \cos\phi \, m(t) \tag{2.11}$$

La señal demodulada $v_o(t)$ es consecuentemente proporcional a $m(t)$ cuando el error de fase ϕ es una constante. La amplitud de esta señal demodulada es máxima cuando $\phi = 0$, y es mínima (cero) cuando $\phi = \pm\pi/2$. La señal demodulada cero, la cual ocurre para $\phi = \pm\pi/2$, representa el *efecto nulo en cuadratura* del detector coherente. De este modo el error de fase ϕ en el oscilador local provoca que la salida del detector sea atenuada por un factor igual a $\cos\phi$. Mientras el error de fase ϕ sea constante, el detector proporciona una versión no distorsionada de la señal de banda base original $m(t)$. Sin embargo, en la práctica solemos encontrar que el error de fase ϕ varía aleatoriamente con el tiempo, debido a variaciones aleatorias en el canal de comunicación. El resultado es que en la salida del detector, el factor multiplicador $\cos\phi$ varía también aleatoriamente con el tiempo, lo cual es obvio que no se desea. Por lo tanto, deben tomarse provisiones en el sistema para mantener el oscilador local en el receptor en perfecto sincronismo, tanto en frecuencia como en fase, con la onda portadora utilizada para generar la señal modulada de DSB-SC en el transmisor. La complejidad del sistema resultante es el precio que debe pagarse al suprimir la onda portadora para ahorrar potencia del transmisor.

■ RECEPTOR DE COSTAS

Un método para obtener un sistema receptor síncrono práctico, adecuado para demodular ondas de DSB-SC, consiste en utilizar el *receptor de Costas*[2] que se muestra en la figura 2.9. Este

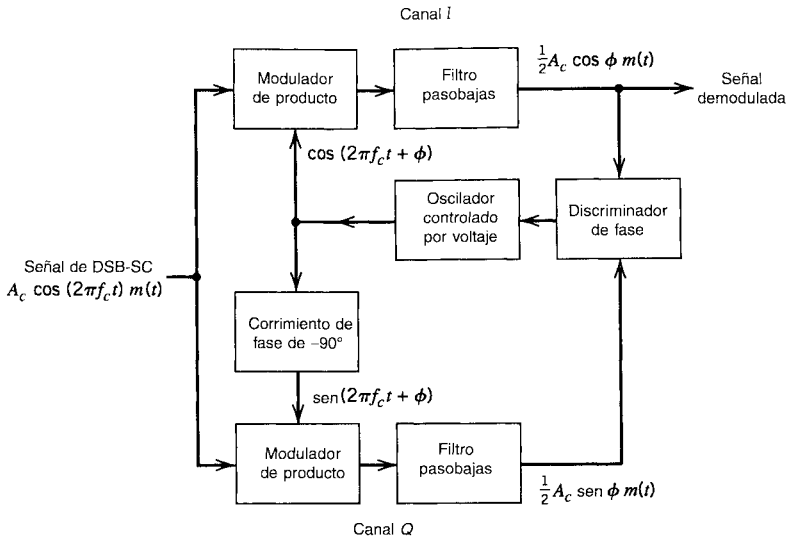


FIGURA 2.9 Receptor de Costas.

receptor está compuesto por dos detectores coherentes alimentados con la misma señal de entrada, a saber la onda entrante de DSB-SC $A_c \cos(2\pi f_c t)m(t)$, pero con señales de oscilador local individuales que están en cuadratura de fase entre sí. La frecuencia del oscilador local se ajusta para que sea la misma que la de la frecuencia de la portadora f_c, la cual se supone que se conoce *a priori*. El detector en la trayectoria superior recibe el nombre de *detector coherente en fase* o *canal I*, y el de la trayectoria inferior se denomina *detector coherente en cuadratura–fase* o *canal Q*. Estos dos detectores se acoplan entre sí para formar un sistema de *realimentación negativa* diseñado de modo que se mantenga síncrono el oscilador local con la onda portadora.

Para entender la operación de este receptor, suponga que la señal del oscilador local es de la misma fase que la onda portadora $A_c \cos(2\pi f_c t)$ utilizada para generar la onda de DSB-SC entrante. En estas condiciones, descubrimos que la salida del canal *I* contiene la señal demodulada $m(t)$ deseada, en tanto que la salida del canal *Q* es cero debido al efecto nulo en cuadratura del canal *Q*. Supóngase a continuación que la fase del oscilador local se desplaza a partir de su valor apropiado en un pequeño ángulo de ϕ radianes. La salida del canal *I* en esencia permanecerá invariable, aunque ahora aparecerá cierta señal en la salida del canal *Q*, que es proporcional a $\text{sen } \phi \simeq \phi$ para ϕ pequeña. Esta salida de cañal *Q* tendrá la misma polaridad que la salida del canal *I* en una dirección del corrimiento de fase del oscilador local y la polaridad opuesta para la dirección opuesta del corrimiento de fase de dicho oscilador. De tal manera, al combinar las salidas del canal *I* y del *Q* en un *discriminador de fase* (el cual consta de un multiplicador seguido por un filtro pasobajas), como se muestra en la figura 2.9, se obtiene una señal de control de cd que corrige de modo automático los errores de fase local en el *oscilador controlado por voltaje*.

Es patente que el control de fase en el receptor de Costas cesa con la modulación y que el cierre de fase tiene que restablecerse con la reaparición de la modulación. Lo anterior no constituye un problema serio cuando se recibe la transmisión de voz, debido a que el proceso de cierre ocurre normalmente tan rápido que no se percibe la distorsión.

■ MULTIPLEXADO DE PORTADORA EN CUADRATURA

El efecto nulo en cuadratura del detector coherente puede aprovecharse en la construcción del llamado *multiplexado de portadora en cuadratura* o *modulación de amplitud por cuadratura* (QAM).

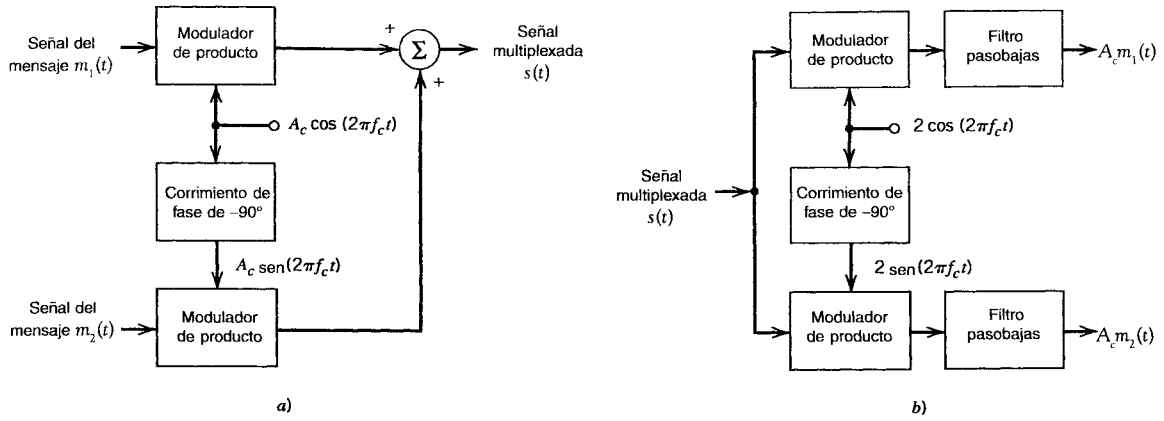


FIGURA 2.10 Sistema de multiplexado de portadora en cuadratura. *a*) Transmisor. *b*) Receptor.

Este esquema permite que dos ondas moduladas de DSB-SC (que resultan de la aplicación de dos señales de mensaje físicamente *independientes*) ocupen el mismo ancho de banda de canal, e incluso posibilita la separación de las dos señales de mensaje en la salida del receptor. Por tanto, resulta un *esquema de conservación de ancho de banda.*

En la figura 2.10 se muestra un diagrama de bloques del sistema de multiplexado de portadora en cuadratura. La parte transmisora del sistema, señalada en la figura 2.10*a*, significa el uso de dos moduladores de producto independientes que se alimentan con dos ondas portadoras de la misma frecuencia pero que difieren en fase en –90 grados. La señal transmitida $s(t)$ consiste en la suma de dos salidas del modulador de producto, como se indica mediante

$$s(t) = A_c m_1(t) \cos(2\pi f_c t) + A_c m_2(t) \operatorname{sen}(2\pi f_c t) \tag{2.12}$$

donde $m_1(t)$ y $m_2(t)$ denotan las dos señales de mensaje diferentes aplicadas a los moduladores de producto. De ese modo $s(t)$ ocupa un ancho de banda de canal de $2W$ centrado en la frecuencia de portadora f_c, donde W es el ancho de banda del mensaje de $m_1(t)$ o $m_2(t)$. De acuerdo con la ecuación (2.12), es posible considerar a $A_c m_1(t)$ como la componente en fase de la señal pasobanda multiplexada $s(t)$ y a $-A_c m_2(t)$ como su componente en cuadratura.

La parte receptora del sistema se muestra en la figura 2.10*b*. La señal multiplexada $s(t)$ se aplica en forma simultánea a dos detectores coherentes independientes que se alimentan con dos portadoras locales de la misma frecuencia pero que difieren en fase en –90 grados. La salida del detector superior es $A_c m_1(t)$, en tanto que la del detector inferior es $A_c m_2(t)$. Para que el sistema opere en forma satisfactoria, es importante mantener la relación entre fase y frecuencia correctas entre los osciladores locales utilizados en las partes del transmisor y del receptor del sistema.

Para mantener esta sincronización, es factible enviar una *señal piloto* fuera de la banda de paso de la señal modulada. En este método, la señal piloto consiste por lo general en un tono senoidal de baja potencia cuya frecuencia y fase se relacionan con la onda portadora $c(t)$; en el receptor, la señal piloto se extrae por medio de un circuito sintonizador adecuado y luego se traslada a la frecuencia correcta para utilizarse en el detector coherente.

■ MODULACIÓN DE BANDA LATERAL ÚNICA

En la modulación de banda lateral única, sólo se transmite la banda lateral superior o inferior. Podemos generar una onda modulada de este tipo utilizando el *método de discriminación de frecuencia* que consta de dos etapas:

- La primera etapa es un modulador de producto, que genera una onda modulada de DSB-SC.
- La segunda etapa es un filtro pasobanda, que se diseña para dejar pasar una de las bandas laterales de esta onda modulada y suprimir la otra.

Desde un punto de vista práctico, el más serio requerimiento de generación de SSB utilizando el método de discriminación de frecuencia surge de la banda lateral indeseada. La componente de frecuencia más cercana de la banda lateral indeseada se separa de la banda lateral deseada duplicando la componente de frecuencia más baja de la señal (moduladora) del mensaje. La implicación aquí es que para que sea posible la generación de una señal modulada de SSB, el espectro del mensaje debe tener una brecha de energía centrada en el origen, como se ilustra en la figura 2.11*a*. Este requerimiento se satisface naturalmente mediante señales de voz, cuya brecha de energía es aproximadamente de 600 Hz de ancho (es decir, se extiende de −300 a +300 Hz). Por tanto, suponiendo que se retiene la banda lateral superior, el espectro de la señal modulada de SSB es como se muestra en la figura 2.11*b*.

Al diseñar el filtro pasobanda utilizado en el discriminador de frecuencia para generar la onda modulada de SSB, debemos satisfacer tres requerimientos básicos:

- La banda lateral deseada se encuentra dentro de la banda de paso del filtro.
- La banda lateral no deseada se ubica dentro de la banda de supresión del filtro.
- La banda de transición del filtro, la cual separa la banda de paso de la banda de supresión, es dos veces la componente de frecuencia más baja de la señal del mensaje.

Este tipo de discriminación de frecuencia suele requerir del empleo de filtros altamente selectivos, los cuales únicamente pueden realizarse en la práctica por medio de resonadores de cristal.

Para demodular una señal modulada de SSB $s(t)$, podemos utilizar un detector coherente, el cual multiplica $s(t)$ por una portadora generada localmente y luego hacer pasar el producto por filtros pasobajas. Este método de demodulación supone sincronismo perfecto entre el oscilador y el detector coherente y el oscilador utilizado para alimentar la onda portadora en el transmisor. Este requerimiento suele satisfacerse de una de dos maneras:

- Una *portadora piloto* de baja potencia se transmite además de la banda lateral elegida.
- Se utiliza un oscilador altamente estable, sintonizado a la misma frecuencia que la de la portadora, en el receptor.

En el último método, irremediablemente existirá algún error de fase ϕ a la salida del oscilador local respecto de la onda portadora que se utiliza para generar la onda modulada de SSB entrante. El efecto de este error de fase es introducir una *distorsión de fase* en la señal demodulada, donde cada componente de frecuencia de la señal del mensaje original sufre un corrimiento de fase constante ϕ.

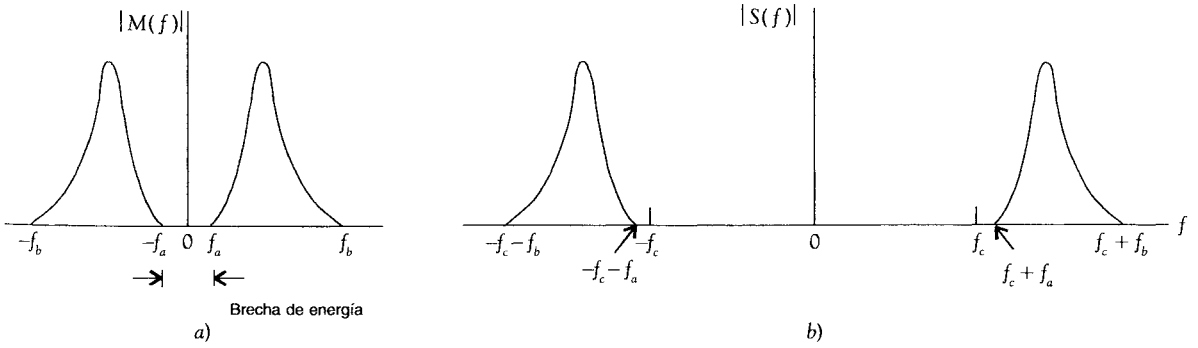


FIGURA 2.11 *a*) Espectro de una señal de mensaje $m(t)$ con una brecha de energía de ancho $2f_a$ centrado en el origen. *b*) Espectro de la señal de SSB correspondiente que contiene la banda lateral superior.

Esta distorsión de fase es tolerable en comunicaciones de voz, debido a que el oído humano resulta relativamente insensible a la distorsión de fase. En particular, la presencia de la distorsión de fase origina un *efecto de voz del Pato Donald*. En la transmisión de señales de música y video, por otro lado, la presencia de esta forma de distorsión de forma de onda es por completo inaceptable.

MODULACIÓN DE BANDA LATERAL RESIDUAL

En la *modulación de banda lateral residual* (VSB), una de las bandas laterales se suprime parcialmente y un residuo de la otra banda lateral se transmite para compensar esa supresión. Un método popular para generar una onda modulada VSB consiste en utilizar el *método de discriminación de frecuencia*. Primero, generamos una onda modulada de DSB-SC y luego la hacemos pasar por un filtro pasobanda, como se muestra en la figura 2.12; es el diseño especial del filtro pasobanda lo que distingue la modulación de VSB de la SSB. Suponiendo que se transmite un residuo de la banda lateral inferior, la respuesta en frecuencia $H(f)$ del filtro pasobanda toma la forma que se muestra en la figura 2.13. Para simplificar las cosas, sólo se muestra aquí la respuesta relativa a las frecuencias positivas. Esta respuesta en frecuencia se normaliza, de modo que la frecuencia de la portadora f_c tiene $|H(f_c)| = 1/2$. La característica importante que debe advertirse en la figura 2.13 es que la porción de corte de la respuesta en frecuencia alrededor de la frecuencia de portadora f_c exhibe *simetría impar*. Esto es, dentro del intervalo de transición $f_c - f_v \leq |f| \leq f_c + f_v$ se satisfacen las siguientes dos condiciones:

1. La suma de los valores de la respuesta en magnitud $|H(f)|$ a cualesquiera dos frecuencias igualmente desplazadas arriba y debajo de f_c es la unidad.
2. El arg $(H(f))$ de la respuesta en fase es lineal. Esto es, $H(f)$ satisface la condición

$$H(f - f_c) + H(f + f_c) = 1 \quad \text{para} - W \leq f \leq W \tag{2.13}$$

Advierta también que fuera de la banda de frecuencia de interés (es decir, $|f| > f_c + W$), la respuesta en frecuencia $H(f)$ quizá tenga una especificación arbitraria. En consecuencia, el ancho de banda de transmisión de la modulación de VSB es

$$B_T = W + f_v \tag{2.14}$$

donde W es el ancho de banda del mensaje y f_v es el ancho de la banda lateral residual.

De acuerdo con la tabla 2.1, la onda modulada de VSB se representa en el dominio del tiempo como

$$s(t) = \frac{1}{2}A_c m(t)\cos(2\pi f_c t) \pm \frac{1}{2}A_c m'(t)\,\text{sen}(2\pi f_c t) \tag{2.15}$$

donde el signo más corresponde a la transmisión de un residuo de la banda lateral superior, y el signo menos a la transmisión de un residuo de la banda lateral inferior. La señal $m'(t)$ en la

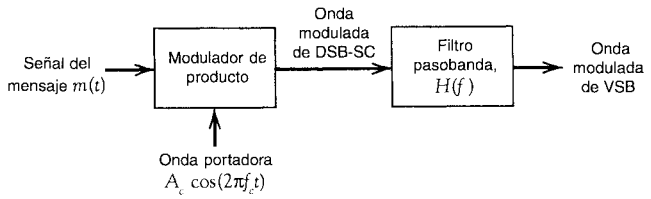


FIGURA 2.12 Esquema de filtrado para la generación de la onda modulada de VSB.

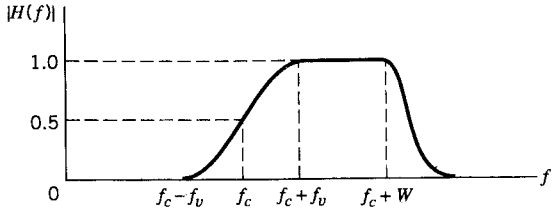


FIGURA 2.13 Respuesta en magnitud del filtro de VSB; sólo se muestra la porción de frecuencia positiva.

componente de cuadratura de $s(t)$ se obtiene al pasar la señal del mensaje $m(t)$ por un filtro cuya respuesta en frecuencia $H_Q(f)$ satisface el siguiente requerimiento (vea el problema 2.20):

$$H_Q(f) = j[H(f - f_c) - H(f + f_c)] \quad \text{para} - W \le f \le W \tag{2.16}$$

La figura 2.14 presenta una gráfica de la respuesta en frecuencia $H_Q(f)$, escalada por $1/j$. El papel de la componente en cuadratura determinada por $H_Q(f)$ es interferir con la componente en fase en la ecuación (2.15) de manera que se reduzca en forma parcial la potencia en una de las bandas laterales de la onda modulada $s(t)$ y se retenga simplemente un residuo de la otra banda lateral, como se desea.

Es interesante advertir que la modulación de SSB puede considerarse como un caso especial de la modulación de VSB. Específicamente, cuando la banda lateral residual se reduce a cero (es decir, dejamos $f_v = 0$), la onda modulada $s(t)$ de la ecuación (2.15) adquiere la forma límite de una onda modulada de banda lateral.

SEÑALES DE TELEVISIÓN

Un análisis de la modulación de banda lateral residual resultaría incompleto sin mencionar su papel en la transmisión de la televisión comercial (TV). Los detalles exactos del formato de modulación utilizado para transmitir la señal de video que caracteriza a un sistema de TV reciben la influencia de dos factores:

1. La señal de video exhibe un gran ancho de banda y un importante contenido de baja frecuencia, lo que sugiere el uso de modulación de banda lateral residual.
2. La circuitería que se usa para la demodulación en el receptor debe ser simple y, en consecuencia, económica; esto sugiere el uso de la detección de envolvente, la cual requiere la adición de una portadora a la onda modulada VSB.

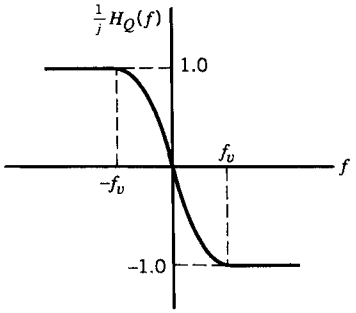


FIGURA 2.14 Respuesta en frecuencia de un filtro para producir la componente en cuadratura de la onda modulada de VSB.

Sin embargo, con respecto al punto 1, debe subrayarse que a pesar de que existe en realidad un deseo fundamental de conservar el ancho de banda, en la transmisión de la TV comercial la señal que se transmite no tiene la suficiente modulación de banda lateral residual. La razón radica en que en el transmisor los niveles de potencias son elevados, con el resultado de que resultaría costoso controlar en forma rígida el filtrado de las bandas laterales. En lugar de esto, se inserta un *filtro* de VSB en cada receptor, donde los niveles de potencia son bajos. El desempeño total es el mismo que el de la modulación convencional de banda lateral residual, salvo por algo de potencia y ancho de banda desperdiciados. Estos señalamientos se ilustran en la figura 2.15. En particular, la figura 2.15*a* presenta el espectro idealizado de la señal de TV transmitida. Se transmiten la banda lateral superior, 25% de la banda lateral inferior y la portadora de la imagen. La respuesta en frecuencia del filtro de VSB utilizado para dar la forma requerida al espectro en el receptor se presenta en la figura 2.15*b*.

El ancho de banda del canal utilizado para la transmisión de TV en Estados Unidos es de 6 MHz, como se señala en la figura 2.15*b*. Este ancho de banda de canal no sólo se ajusta al requerimiento de ancho de banda de la señal de video modulada de VSB, sino que también proporciona la señal de sonido acompañante que modula una portadora propia. Los valores presentados en el eje de frecuencia en las figuras 2.15*a* y 2.15*b* pertenecen a un canal de TV específico. De acuerdo con esta figura, la frecuencia de la portadora de la imagen corresponde a 55.25 MHz, y la frecuencia de la portadora del sonido a 59.75 MHz. Sin embargo, advierta que el contenido de información de la señal de TV yace en el *espectro de banda base* que se extiende desde 1.25 MHz debajo de la portadora de la imagen hasta 4.5 MHz sobre ella.

En lo que se refiere al punto 2, el empleo de la detección de envolventes (aplicada a la onda modulada de VSB más la portadora) produce una *distorsión de la forma de onda* en la señal

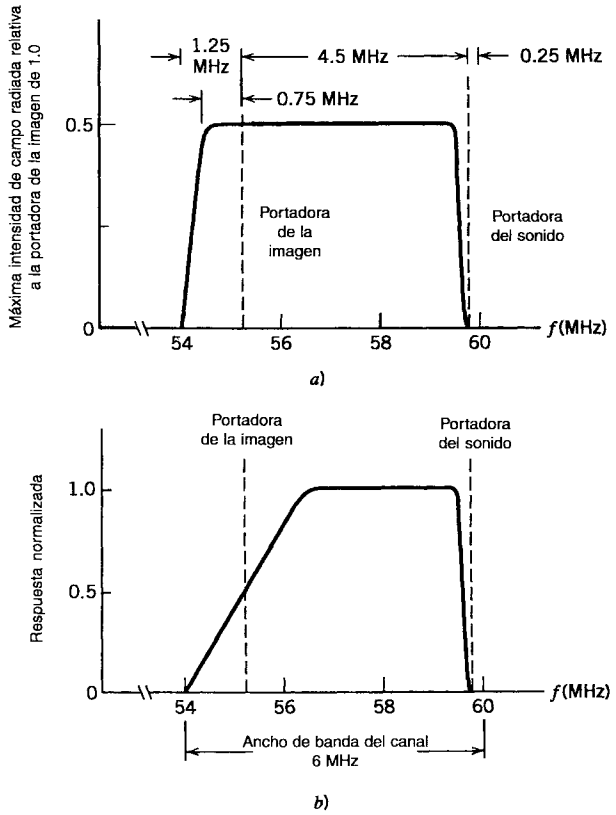


FIGURA 2.15 *a*) Espectro de magnitud idealizado de una señal de TV transmitida. *b*) Respuesta en magnitud del filtro de formación de VSB en el receptor.

de video que se recupera en la salida del detector. La distorsión la produce la componente en cuadratura de la onda modulada de VSB; este aspecto se explica a continuación.

El uso de la descripción en el dominio del tiempo que se da en la ecuación (2.15) permite la determinación de la distorsión de la forma de onda que provoca el detector de envolvente. De modo específico, al sumar la componente de la portadora $A_c \cos(2\pi f_c t)$ a la onda modulada de VSB de la ecuación (2.15), siendo escalada la última por un factor k_a se modifica la señal aplicada k_a a la entrada del detector de envolvente como

$$s(t) = A_c\left[1+\frac{1}{2}k_a m(t)\right]\cos(2\pi f_c t) \pm \frac{1}{2}k_a A_c m'(t)\,\text{sen}(2\pi f_c t) \tag{2.17}$$

donde la constante k_a determina la modulación porcentual. La salida del detector de envolvente, denotada por $a(t)$, es entonces

$$\begin{aligned} a(t) &= A_c\left\{\left[1+\frac{1}{2}k_a m(t)\right]^2 + \left[\frac{1}{2}k_a m'(t)\right]^2\right\}^{1/2} \\ &= A_c\left[1+\frac{1}{2}k_a m(t)\right]\left\{1+\left[\frac{\frac{1}{2}k_a m'(t)}{1+\frac{1}{2}k_a m(t)}\right]^2\right\}^{1/2} \end{aligned} \tag{2.18}$$

La ecuación (2.18) indica que a la distorsión se agrega $m'(t)$, que es la responsable de la componente en cuadratura de la señal modulada de VSB de entrada. Dicha distorsión se puede reducir recurriendo a dos métodos:

- Reduciendo el porcentaje de modulación para disminuir la sensibilidad de amplitud k_a.
- Aumentando el ancho de la banda lateral residual para reducir $m'(t)$.

De hecho, en la práctica se emplean ambos métodos. En la transmisión de TV comercial, el ancho de la banda lateral residual, que es más o menos de 0.75 MHz o un sexto de la banda lateral completa, se determina para mantener la distorsión debida a $m'(t)$ dentro de límites tolerables cuando la modulación porcentual es cerca de 100.

2.4 *Traslación de frecuencia*

La operación básica implicada en la modulación de una banda lateral es de hecho una forma de *traslación de frecuencia*, que es la causa de que algunas veces la modulación de banda lateral se denomine *cambio de frecuencia, mezcla* o *heterodinación*. Esta operación se ilustra con claridad en el espectro de la señal de la figura 2.11*b* en comparación con la señal del mensaje original en la figura 2.11*a*. De manera específica, vemos que el espectro del mensaje que ocupa la banda de f_a a f_b para frecuencias positivas en la figura 2.11*a* se desplaza hacia arriba en una cantidad igual a la frecuencia de la portadora f_c en la figura 2.11*b*, y que el espectro del mensaje para frecuencias negativas se traslada hacia abajo de una manera simétrica.

Es factible generalizar del siguiente modo la idea de traslación de frecuencia que acabamos de describir. Suponga que tenemos una onda modulada $s_1(t)$ cuyo espectro se centra en la frecuencia de portadora f_1, y que el requerimiento es trasladarla hacia frecuencias superiores de manera que la frecuencia de portadora cambie de f_1 a un nuevo valor f_2. Es posible satisfacer este requerimiento utilizando el *mezclador* que se muestra en la figura 2.16. Específicamente, el *mezclador* es un dispositivo que consiste en un modulador de producto seguido por un filtro pasobanda.

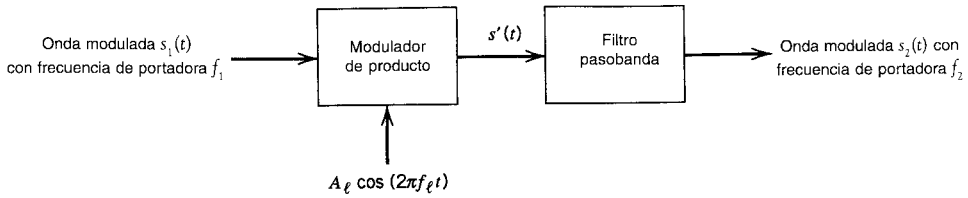


FIGURA 2.16 Diagrama de bloques del mezclador.

Para explicar la acción del mezclador, considere la situación representada en la figura 2.17, donde, con fines ilustrativos, se supone que la entrada del mezclador $s_1(t)$ es una señal de AM con frecuencia de portadora f_1 y ancho de banda $2W$. La parte *a* de la figura 2.17 presenta el espectro $S_1(f)$ de AM al suponer que $f_1 > W$. La parte *b* de la figura presenta el espectro $S'(f)$ de la señal resultante $s'(t)$ a la salida del modulador de producto.

La señal $s'(t)$ puede considerarse como la suma de dos componentes moduladas: una componente representada por el espectro sombreado de la figura 2.17*b*, y la otra componente representada por el espectro sin sombra en esta figura. Dependiendo de si la frecuencia entrante de portadora f_1 se traslade hacia arriba o hacia abajo, es posible identificar dos situaciones diferentes, según se describe aquí:

Conversión ascendente. En este caso la frecuencia de portadora trasladada f_2 es mayor que la frecuencia de portadora entrante f_1 y, por tanto, la frecuencia f_l del oscilador local que se requiere está definida por

$$f_2 = f_1 + f_l$$

o bien

$$f_l = f_2 - f_1$$

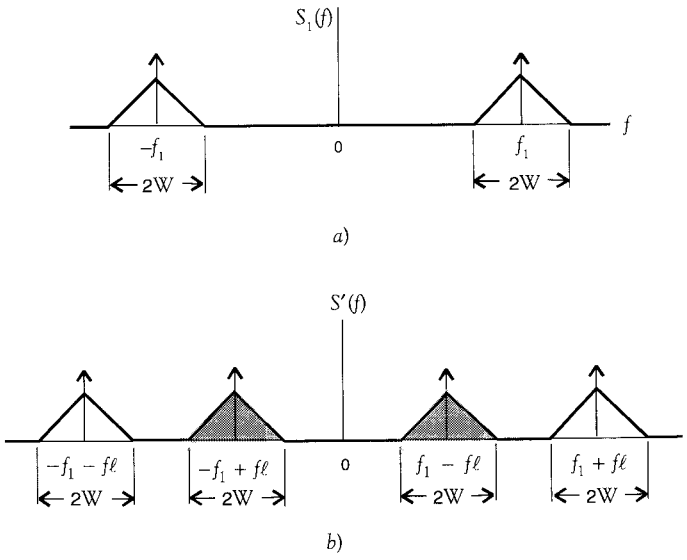


FIGURA 2.17 *a*) Espectro de la señal modulada $s_1(t)$ a la entrada del mezclador. *b*) Espectro de la señal correspondiente $s'(t)$ a la salida del modulador de producto en el mezclador.

La parte sin sombra del espectro de la figura 2.17*b* define la señal modulada deseada $s_2(t)$, y la parte sombreada de este espectro define la *señal de imagen* asociada con $s_2(t)$. Por razones obvias, el mezclador en este caso se conoce como un *convertidor ascendente de frecuencia.*

Conversión descendente. En este segundo caso la frecuencia de portadora trasladada f_2 es más pequeña que la frecuencia de portadora entrada f_1, y la frecuencia f_l que se requiere del oscilador está, por tanto, definida por

$$f_2 = f_1 - f_l$$

o bien

$$f_l = f_1 - f_2$$

La imagen que tenemos esta vez es la inversa de la que corresponde a la conversión ascendente. En particular, la parte sombreada del espectro en la figura 2.17*b* define la señal modulada que se desea $s_2(t)$, y la parte sin sombra de este espectro determina la señal de la imagen asociada. El mezclador se conoce en este caso como un *convertidor descendente de frecuencia.* Advierta que en esas condiciones la frecuencia de portadora trasladada f_2 tiene que ser mayor que W (es decir, la mitad del ancho de banda de la señal modulada) para evitar el traslape de banda lateral.

El propósito del filtro pasobanda en el mezclador de la figura 2.16 es dejar pasar la señal modulada deseada $s_2(t)$ y eliminar la señal asociada de la imagen. Lo anterior se consigue alineando la frecuencia de banda media del filtro con la frecuencia de la portadora trasladada f_2 y asignándole un ancho de banda igual al de la señal modulada entrante $s_1(t)$.

Es importante advertir que la mezcla es una operación *lineal.* En consecuencia, se preserva por completo la relación de las bandas laterales de la onda modulada entrante con la portadora a la salida del mezclador.

2.5 *Multiplexado por división de frecuencia*

Otra operación de procesamiento de señales importante es el *multiplexado*; mediante éste es posible combinar varias señales independientes en una señal compuesta adecuada para la transmisión por un canal común. Las frecuencias de voz que transmiten los sistemas telefónicos, por ejemplo, van de 300 a 3100 Hz. Para transmitir algunas de estas señales por el mismo canal, dichas señales deben mantenerse apartadas de modo que no interfieran entre sí y para que de ese modo sea posible separarlas en el extremo de recepción. Lo anterior se logra al separar las señales en la frecuencia o en el tiempo. La técnica de separar las señales en la frecuencia se conoce como *multiplexado por división de frecuencia* (FDM), en tanto que la técnica de separar las señales en el tiempo se denomina *multiplexado por división de tiempo* (TDM). En esta sección explicamos los sistemas FDM y los TDM se exponen en el capítulo 3.

En la figura 2.18 se presenta un diagrama de bloques de un sistema FDM. Se supone que las señales del mensaje entrante son del tipo pasobajas, aunque su espectro no necesariamente tiene valores distintos de cero en todo momento mientras tiende a la frecuencia cero. Luego de cada señal de entrada, mostramos un filtro pasobajas, el cual se diseña para eliminar las componentes de alta frecuencia que no contribuyen de manera significativa a la representación de la señal, pero que son capaces de perturbar otras señales de mensaje que comparten el canal común. Estos filtros pasobajas pueden omitirse sólo si las señales de entrada están lo suficientemente limitadas de banda en un principio. Las señales filtradas se aplican en moduladores que corren los intervalos de frecuencia de las señales hasta ocupar intervalos de frecuencia exclusivamente

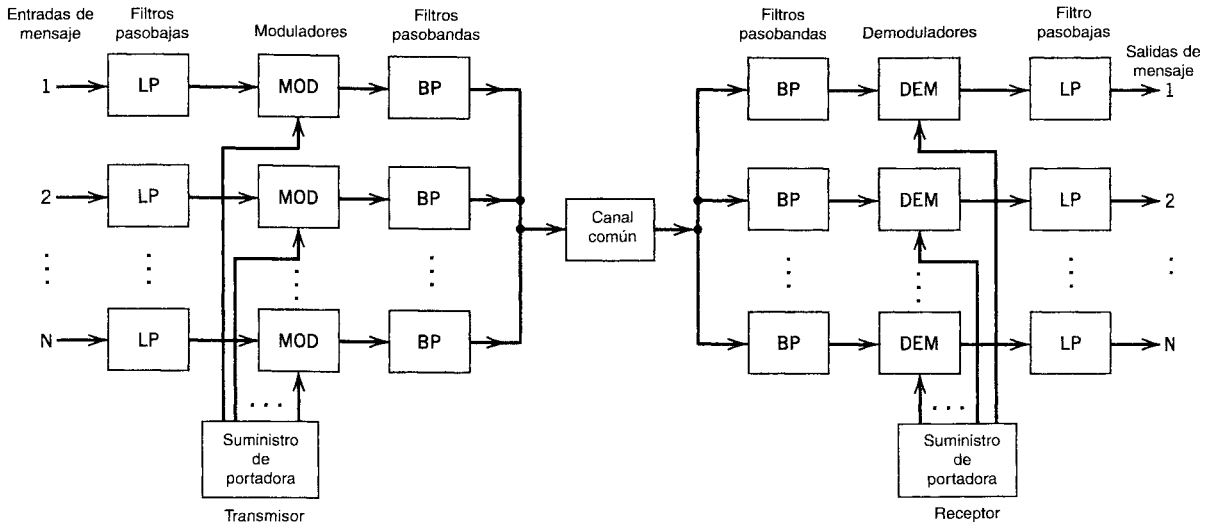


FIGURA 2.18 Diagrama de bloques del sistema de FDM.

mutuos. Las frecuencias de portadora necesarias para efectuar estas traslaciones de frecuencia se obtienen de un suministro de portadoras. Para la modulación, es posible recurrir a cualquiera de los métodos que se describieron en las secciones anteriores de este capítulo. Sin embargo, el que más se utiliza en la modulación del multiplexado por división de frecuencia es el método de banda lateral, que, en el caso de señales de voz, requiere de un ancho de banda aproximadamente igual al de la señal de voz original. En la práctica, a cada entrada de voz suele asignársele un ancho de banda de 4 kHz. Los filtros pasobanda que siguen a los moduladores se utilizan para restringir la banda de cada onda modulada en su intervalo preestablecido. Las salidas que resultan del filtro pasobanda se combinan luego en paralelo para formar la entrada al canal común. En la terminal de recepción se utiliza un banco de filtros pasobanda, con sus entradas conectadas en paralelo, para separar las señales de mensaje en un esquema de ocupación de frecuencia. Por último, las señales de mensaje originales se recuperan por medio de demoduladores individuales. Advierta que el sistema FDM que se presenta en la figura 2.18 opera sólo en una dirección. Para proporcionar la transmisión de dos vías, como en la telefonía, tenemos que duplicar por completo los equipos de multiplexado, con los componentes conectados en orden inverso y con ondas de la señal que procedan de derecha a izquierda.

EJEMPLO 2.1

La implantación práctica de un sistema FDM suele significar muchos pasos de modulación y demodulación, como se ilustra en la figura 2.19. El primer paso de multiplexado combina 12 entradas de voz en un *grupo básico*, que se forma cuando la n-ésima entrada modulada tiene una portadora a frecuencia $f_c = 60 + 4n$ kHz, donde $n = 1, 2, \ldots, 12$. Después se eligen las bandas laterales inferiores mediante un filtrado pasobanda y se combinan para formar un grupo de 12 bandas laterales inferiores (una para cada entrada de voz). De esa manera, el grupo básico ocupa la banda de frecuencia de 60 a 108 kHz. El siguiente paso en la jerarquía FDM significa la combinación de cinco grupos básicos en un *supergrupo*. Esto se consigue al utilizar el grupo n-ésimo para modular una portadora de frecuencia $f_c = 372 + 48n$ kHz, donde $n = 1, 2, \ldots, 5$. De nuevo aquí las bandas laterales inferiores se eligen por medio de filtrado y luego se combinan para formar un supergrupo que ocupa la banda de 312 a 552 kHz. De este modo se diseña un supergrupo para aceptar 60

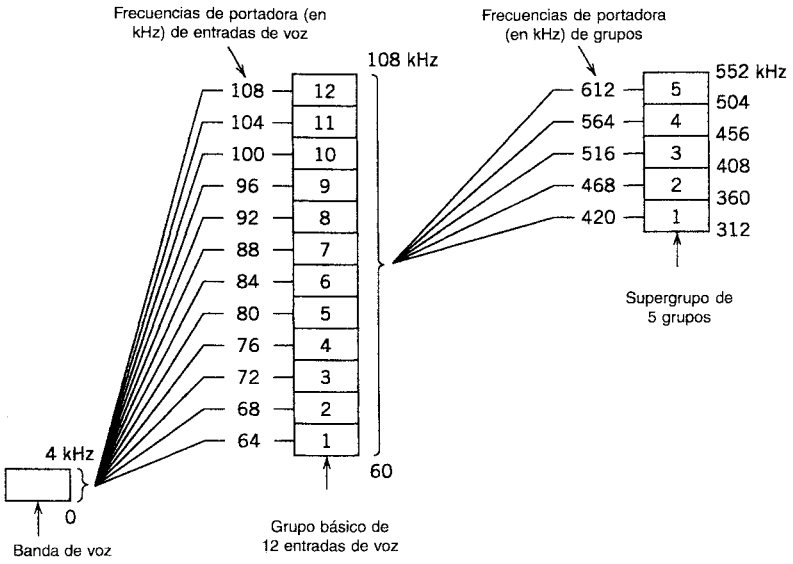


FIGURA 2.19 Ilustración de las etapas de modulación en un sistema de FDM.

entradas de voz independientes. La razón por la que se forma de este modo el supergrupo es que los filtros económicos de las características requeridas se obtienen sólo para un intervalo limitado de frecuencia. De manera similar, los supergrupos se combinan en *grupos maestros*, y éstos en *grupos sumamente grandes*.

2.6 *Modulación angular*

En las secciones previas de este capítulo investigamos el efecto de variar lentamente la amplitud de una onda portadora senoidal de acuerdo con la señal de la banda base (la que contiene la información). Existe otro modo de modular una onda portadora senoidal, a saber, la *modulación angular*, en la cual se varía el ángulo de la onda portadora de acuerdo con la señal de la banda base. En este método de modulación se mantiene constante la amplitud de la onda portadora. Una característica importante de la modulación angular es que puede ofrecer una mejor discriminación con respecto al ruido y la interferencia que la modulación de amplitud. Sin embargo, como se demostrará después en la sección 2.7, esta mejora en el desempeño se consigue a costa de un ancho de banda de transmisión incrementado; esto es, la modulación angular nos proporciona medios prácticos para intercambiar ancho de banda de canal por un mejor desempeño relativo al ruido. Un compromiso de este tipo *no* es posible con la modulación de amplitud, sin que importe su forma.

■ DEFINICIONES BÁSICAS

Dejemos que $\theta_i(t)$ denote el *ángulo* de la portadora senoidal modulada, que se supone es la función de la señal del mensaje. Expresamos la *onda modulada* en ángulo que resulta como

$$s(t) = A_c \cos[\theta_i(t)] \tag{2.19}$$

donde A_c es la amplitud de la portadora. Ocurre una oscilación completa siempre que $\theta_i(t)$ cambie en 2π radianes. Si $\theta_i(t)$ aumenta monotónicamente con el tiempo, la frecuencia promedio en hertz, sobre un intervalo de t a $t + \Delta t$, está dada por

$$f_{\Delta t}(t) = \frac{\theta_i(t + \Delta t) - \theta_i(t)}{2\pi\,\Delta t} \tag{2.20}$$

De este modo, es posible definir la *frecuencia instantánea* de la señal modulada en ángulo $s(t)$ de la siguiente manera:

$$\begin{aligned} f_i(t) &= \lim_{\Delta t \to 0} f_{\Delta t}(t) \\ &= \lim_{\Delta t \to 0}\left[\frac{\theta_i(t + \Delta t) - \theta_i(t)}{2\pi\,\Delta t}\right] \\ &= \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \end{aligned} \tag{2.21}$$

Por tanto, según la ecuación (2.19), podemos interpretar la señal modulada en ángulo $s(t)$ como un fasor rotatorio de longitud A_c y ángulo $\theta_i(t)$. La velocidad angular de un fasor de este tipo es $d\theta_i(t)/dt$ medida en radianes por segundo, de acuerdo con la ecuación (2.21). En el caso simple de una portadora no modulada, el ángulo $\theta_i(t)$ es

$$\theta_i(t) = 2\pi f_c t + \phi_c$$

y el fasor correspondiente gira con una velocidad angular constante igual a $2\pi f_c$. La constante ϕ_c es el valor de $\theta_i(t)$ en $t = 0$.

Hay un número infinito de formas en las cuales es posible variar el ángulo $\theta_i(t)$ de alguna manera con la señal del mensaje (banda base). Sin embargo, sólo consideraremos dos métodos que suelen usarse, la modulación de fase y la modulación de frecuencia, definidas como sigue:

1. La *modulación de fase* (PM) *es aquella forma de modulación angular en la que el ángulo* $\theta_i(t)$ *se varía linealmente con la señal del mensaje* $m(t)$ de acuerdo con

$$\theta_i(t) = 2\pi f_c t + k_p m(t) \tag{2.22}$$

El término $2\pi f_c t$ representa el ángulo de la portadora *no modulada*; y la constante k_p la *sensibilidad de fase* del modulador, expresada en radianes por volt con la suposición de que $m(t)$ es una forma de onda de voltaje. Por conveniencia, hemos supuesto en la ecuación (2.22) que el ángulo de la portadora no modulada es cero en $t = 0$. La señal modulada en fase $s(t)$ se describe entonces en el dominio del tiempo por medio de

$$s(t) = A_c \cos\left[2\pi f_c t + k_p m(t)\right] \tag{2.23}$$

2. La *modulación de frecuencia* (FM) *es aquella forma de modulación angular en la que la frecuencia instantánea* $f_i(t)$ *se varía linealmente con la señal del mensaje* $m(t)$, como se muestra por medio de

$$f_i(t) = f_c + k_f m(t) \tag{2.24}$$

El término f_c representa la frecuencia de la portadora no modulada, y la constante k_f, la *sensibilidad de frecuencia* del modulador, expresada en hertz por volt suponiendo que $m(t)$

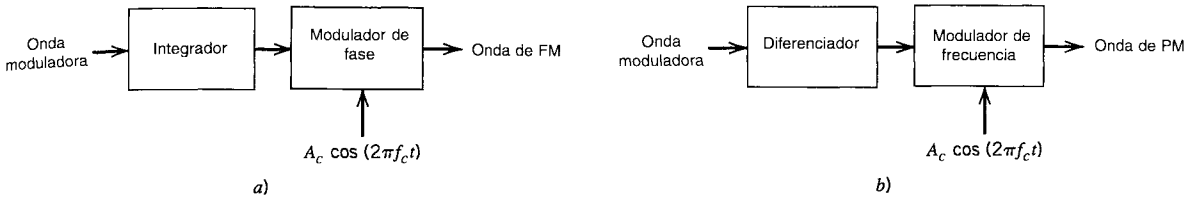


FIGURA 2.20 Ilustración de la relación entre la modulación de frecuencia y la modulación de fase. *a*) Esquema para generar una onda de FM utilizando un modulador de fase. *b*) Esquema para generar una onda de PM utilizando un modulador de frecuencia.

es una forma de onda de voltaje. Al integrar la ecuación (2.24) con respecto al tiempo y multiplicando el resultado por 2π, obtenemos

$$\theta_i(t) = 2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\, d\tau \tag{2.25}$$

donde, por conveniencia, se ha supuesto que el ángulo de la onda portadora no modulada es cero en $t = 0$. La señal modulada en frecuencia se describe, por tanto, en el dominio del tiempo mediante

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\, d\tau\right] \tag{2.26}$$

Una consecuencia por permitir que el ángulo $\theta_i(t)$ se vuelva dependiente de la señal del mensaje $m(t)$, como en la ecuación (2.22), o de su integral, como en la ecuación (2.25), es que los cruces por cero de la señal PM o de la señal FM ya no tienen una regularidad perfecta en su espaciamiento; los cruces por cero se refieren a los instantes en los cuales la forma de onda cambia de un valor negativo a positivo o viceversa. Lo anterior constituye un rasgo importante que distingue a las señales tanto PM como FM de una señal de AM. Otra diferencia importante es que la envolvente de una señal PM o FM es constante (igual a la amplitud de la portadora), en tanto que la envolvente de una señal de AM depende de la señal del mensaje.

La comparación de las ecuaciones (2.23) y (2.26) revela que una señal de FM puede considerarse como una de PM en la que la onda moduladora es $\int_0^t m(\tau)\, d\tau$ en lugar de $m(t)$. Esto quiere decir que una señal de FM puede generarse al integrar primero $m(t)$ y utilizar después el resultado como la entrada a un modulador de fase (ver figura 2.20*a*). De manera inversa, es posible generar una señal de PM al diferenciar primero $m(t)$ y emplear luego el resultado como la entrada a un modulador de frecuencia (ver figura 2.20*b*). De este modo, resulta factible que deduzcamos todas las propiedades de las señales PM de aquellas de las señales de FM y viceversa. De aquí en adelante concentraremos nuestra atención en las señales de FM.

2.7 *Modulación de frecuencia*

La señal de FM $s(t)$ definida por la ecuación (2.26) es una función no lineal de la señal moduladora $m(t)$, la cual hace que la modulación de frecuencia sea un *proceso de modulación no lineal*. En consecuencia, a diferencia de la modulación de amplitud, el espectro de una señal de FM no se relaciona de manera simple con el de la señal moduladora; en vez de eso, su análisis es mucho más difícil que el de una señal de AM.

¿Cómo podemos entonces abordar el análisis espectral de una señal de FM? Proponemos dar una respuesta empírica a esta importante pregunta procediendo del modo siguiente:

- Consideramos el caso más simple posible; es decir, el de una modulación de un tono que produce una señal de FM de banda angosta.
- Consideramos después el caso más general incluyendo también la modulación de un tono, pero esta vez la señal de FM es de banda ancha.

Desde luego, existe la posibilidad de que prosigamos y consideremos el caso más elaborado de una señal de FM multitono. Sin embargo, proponemos no hacerlo, ya que nuestro objetivo inmediato es establecer una relación empírica entre el ancho de banda de transmisión de una señal de FM y el ancho de banda del mensaje. Como veremos más adelante, el análisis espectral de dos etapas que se describe aquí nos ofrece suficiente información para proponer una solución al problema.

Considere entonces una señal moduladora senoidal definida por

$$m(t) = A_m \cos(2\pi f_m t) \tag{2.27}$$

La frecuencia instantánea de la señal de FM que se produce es igual a

$$\begin{aligned} f_i(t) &= f_c + k_f A_m \cos(2\pi f_m t) \\ &= f_c + \Delta f \cos(2\pi f_m t) \end{aligned} \tag{2.28}$$

donde

$$\Delta f = k_f A_m \tag{2.29}$$

La cantidad Δf se denomina la *desviación de frecuencia*, la cual representa la máxima desviación de la frecuencia instantánea de la señal de FM respecto a la frecuencia de la portadora f_c. Una característica fundamental de una señal de FM es que la desviación de frecuencia Δf es proporcional a la amplitud de la señal moduladora e independiente de la frecuencia de modulación.

Recurriendo a la ecuación (2.28), el ángulo $\theta_i(t)$ de la señal de FM se obtiene como

$$\begin{aligned} \theta_i(t) &= 2\pi \int_0^t f_i(\tau)\, d\tau \\ &= 2\pi f_c t + \frac{\Delta f}{f_m} \operatorname{sen}(2\pi f_m t) \end{aligned} \tag{2.30}$$

La proporción entre la desviación de frecuencia Δf y la frecuencia de modulación f_m suele denominarse el *índice de modulación* de la señal de FM. La denotamos por medio de β, y así escribimos

$$\beta = \frac{\Delta f}{f_m} \tag{2.31}$$

y

$$\theta_i(t) = 2\pi f_c t + \beta \operatorname{sen}(2\pi f_m t) \tag{2.32}$$

De acuerdo con la ecuación (2.32) observamos que, en un sentido físico, el parámetro β representa la desviación de fase de la señal FM, esto es, la desviación máxima del ángulo $\theta_i(t)$ con respecto al ángulo $2\pi f_c t$ de la portadora no modulada. Por tanto, β se mide en radianes.

La propia señal de FM está dada por

$$s(t) = A_c \cos[2\pi f_c t + \beta \operatorname{sen}(2\pi f_m t)] \tag{2.33}$$

Dependiendo del valor del índice de modulación β, es posible que distingamos dos casos de modulación en frecuencia:

- FM de *banda angosta*, para la cual β es pequeña comparada con un radián.
- FM de *banda ancha*, para la cual β es grande en comparación con un radián.

Estos dos casos se consideran a continuación, en ese orden.

Modulación de frecuencia de banda angosta

Considere la ecuación (2.33), la cual define una señal de FM que resulta del uso de una señal moduladora senoidal. Al expandir esta relación, obtenemos

$$s(t) = A_c \cos(2\pi f_c t) \cos[\beta \operatorname{sen}(2\pi f_m t)] - A_c \operatorname{sen}(2\pi f_c t) \operatorname{sen}[\beta \operatorname{sen}(2\pi f_m t)] \tag{2.34}$$

Suponiendo que el índice de modulación β es pequeño comparado con un radián, es posible utilizar las siguientes aproximaciones:

$$\cos[\beta \operatorname{sen}(2\pi f_m t)] \simeq 1$$

y

$$\operatorname{sen}[\beta \operatorname{sen}(2\pi f_m t)] \simeq \beta \operatorname{sen}(2\pi f_m t)$$

Por tanto, la ecuación (2.34) se simplifica en

$$s(t) \quad A_c \cos(2\pi f_c t) - \beta A_c \operatorname{sen}(2\pi f_c t) \operatorname{sen}(2\pi f_m t) \tag{2.35}$$

La ecuación (2.35) define la forma aproximada de una señal de FM de banda angosta que produce una señal moduladora senoidal $A_m \cos(2\pi f_m t)$. De esta representación deducimos el modulador que se muestra en forma de diagrama de bloques en la figura 2.21. Este modulador incluye la división de la onda portadora $A_c \cos(2\pi f_c t)$ en dos trayectorias. Una de ellas es directa; la otra trayectoria contiene una red de corrimiento de fase de –90 grados y un modulador de producto, cuya combinación genera una señal modulada de DSB-SC. La diferencia entre estas dos señales produce una señal de FM de banda angosta, pero con cierta distorsión.

En el caso ideal, una señal de FM tiene una envolvente constante y, para el caso de una señal moduladora senoidal de frecuencia f_m, el ángulo $\theta_i(t)$ también es una senoidal con la

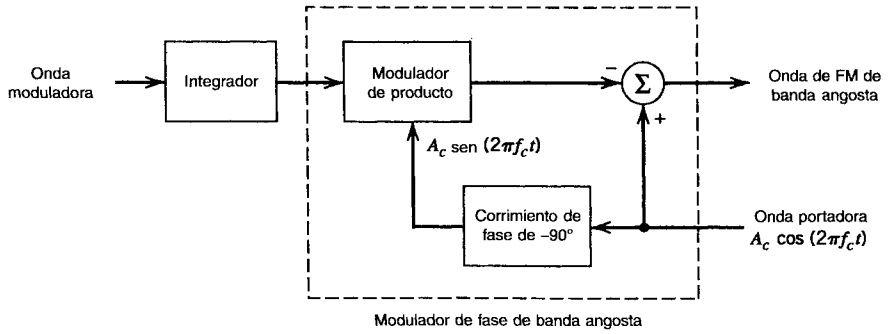


FIGURA 2.21 Diagrama de bloques de un método para generar una señal de FM de banda angosta.

misma frecuencia. Sin embargo, la señal que produce el modulador de banda angosta de la figura 2.21 difiere de esta condición ideal en dos aspectos fundamentales:

1. La envolvente contiene una modulación de amplitud *residual* y, por tanto, varía con el tiempo.
2. Para una onda moduladora senoidal, el ángulo $\theta_i(t)$ contiene *distorsión armónica* en la forma de armónicas de tercer orden y orden superior de la frecuencia de modulación f_m.

No obstante, restringiendo el índice de modulación en $\beta \leq 0.3$ radianes, los efectos de la AM residual y de la PM armónica se limitan a niveles que es posible ignorar.

Regresando a la ecuación (2.35), resulta factible expandirla como sigue:

$$s(t) \simeq A_c \cos(2\pi f_c t) + \frac{1}{2}\beta A_c \{\cos[2\pi(f_c + f_m)t] - \cos[2\pi(f_c - f_m)t]\} \tag{2.36}$$

Esta expresión es similar a la correspondiente que define a la señal de AM, la cual es del siguiente modo:

$$s_{AM}(t) = A_c \cos(2\pi f_c t) + \frac{1}{2}\mu A_c \{\cos[2\pi(f_c + f_m)t] + \cos[2\pi(f_c - f_m)t]\} \tag{2.37}$$

donde μ es el factor de modulación de la señal AM. Al comparar las ecuaciones (2.36) y (2.37) vemos que en el caso de la modulación senoidal, la diferencia básica entre una señal de AM y una señal de FM de banda angosta es que se invierte el signo algebraico de la frecuencia del lado inferior en la FM de banda angosta. De esa manera, una señal de FM de banda angosta requiere esencialmente el mismo ancho de banda de transmisión (es decir, $2f_m$) que la señal de AM.

Podemos representar la señal de FM de banda angosta con un diagrama de fasores en la forma que se indica en la figura 2.22*a*, donde se ha usado como referencia el fasor de la portado-

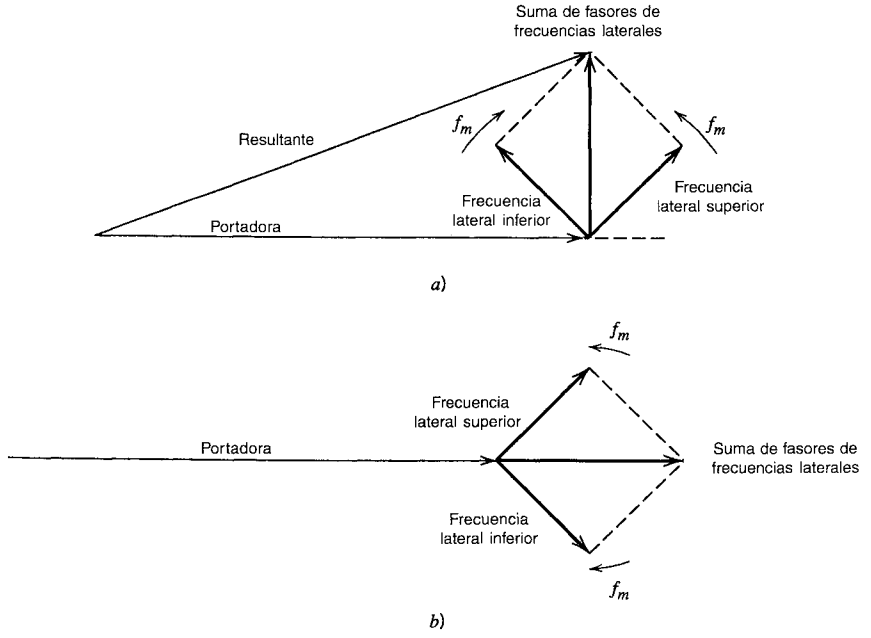


FIGURA 2.22 Una comparación fasorial de las ondas de FM y AM de banda angosta para modulación senoidal. *a*) Onda de FM de banda angosta. *b*) Onda de AM.

ra. Vemos que la resultante de los fasores de las frecuencias de los dos lados siempre está en ángulo recto con el fasor de la portadora. El efecto de esto es producir un fasor resultante que representa a la señal de FM de banda angosta, que es aproximadamente de la misma amplitud que el fasor de la portadora, aunque fuera de fase con respecto a ella. Este diagrama fasorial debe compararse con el de la figura 2.22*b*, que representa una señal de AM. En este último caso vemos que el fasor resultante que representa a la señal de AM tiene una amplitud diferente del fasor de la portadora, aunque siempre en fase con él.

Modulación de frecuencia de banda ancha

Nos interesa determinar ahora el espectro de la señal de FM de un tono relativo a la ecuación (2.33) para un valor arbitrario del índice de modulación β. En general, una señal de FM producida por una señal moduladora senoidal, como en la ecuación (2.33), es en sí misma no periódica a menos que la frecuencia de la portadora f_c sea un múltiplo integral de la frecuencia de modulación f_m. Sin embargo, es posible simplificar las cosas utilizando la representación compleja de señales de la banda de paso que se describe en el apéndice 2. De manera específica, suponemos que la frecuencia de la portadora f_c es lo bastante grande (en comparación con el ancho de banda de la señal de FM) para justificar la reescritura de esta ecuación en la forma

$$\begin{aligned} s(t) &= \mathrm{Re}[A_c \exp(j2\pi f_c t + j\beta \operatorname{sen}(2\pi f_m t))] \\ &= \mathrm{Re}[\tilde{s}(t)\exp(j2\pi f_c t)] \end{aligned} \tag{2.38}$$

donde $\tilde{s}(t)$ es la *envolvente compleja* de la señal de FM $s(t)$, definida por

$$\tilde{s}(t) = A_c \exp[j\beta \operatorname{sen}(2\pi f_m t)] \tag{2.39}$$

De tal modo, a diferencia de la señal de FM original $s(t)$, la envolvente compleja $\tilde{s}(t)$ es una función periódica en el tiempo con una frecuencia fundamental igual a la frecuencia de modulación f_m. Por tanto, podemos expandir $\tilde{s}(t)$ en la forma de una serie compleja de Fourier de la manera siguiente:

$$\tilde{s}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_m t) \tag{2.40}$$

donde el coeficiente complejo de Fourier c_n se define como

$$\begin{aligned} c_n &= f_m \int_{-1/2f_m}^{1/2f_m} \tilde{s}(t)\exp(-j2\pi n f_m t)\, dt \\ &= f_m A_c \int_{-1/2f_m}^{1/2f_m} \exp[j\beta \operatorname{sen}(2\pi f_m t) - j2\pi n f_m t]\, dt \end{aligned} \tag{2.41}$$

Definimos una nueva variable:

$$x = 2\pi f_m t \tag{2.42}$$

Por tanto, es factible que volvamos a escribir la ecuación (2.41) en la nueva forma

$$c_n = \frac{A_c}{2\pi} \int_{-\pi}^{\pi} \exp[j(\beta \operatorname{sen} x - nx)]\, dx \tag{2.43}$$

La integral en el segundo miembro de la ecuación (2.43), salvo por el factor de escalamiento, se reconoce como la *función de Bessel de orden n-ésimo del primer tipo*[3] y argumento β. Esta función se denota comúnmente por medio del símbolo $J_n(\beta)$, como indica

$$J_n(\beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp[j(\beta \operatorname{sen} x - nx)]\, dx \tag{2.44}$$

En consecuencia, es posible reducir la ecuación (2.43) en

$$c_n = A_c J_n(\beta) \tag{2.45}$$

Al sustituir la ecuación (2.45) en la (2.40) obtenemos, en términos de la función de Bessel $J_n(\beta)$, la siguiente expansión para la envolvente compleja de la señal de FM:

$$\tilde{s}(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \exp(j2\pi n f_m t) \tag{2.46}$$

A continuación, mediante la sustitución de la ecuación (2.46) en la (2.38), obtenemos

$$s(t) = A_c \cdot \mathrm{Re}\left[\sum_{n=-\infty}^{\infty} J_n(\beta) \exp[j2\pi(f_c + nf_m)t]\right] \tag{2.47}$$

Con el intercambio del orden de la sumatoria y la evaluación de la parte real en el lado derecho de la ecuación (2.47), llegamos finalmente a

$$s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + nf_m)t] \tag{2.48}$$

Ésta es la forma deseada para la representación en series de Fourier en la señal de FM de un tono $s(t)$ para un valor arbitrario de β. El espectro discreto de $s(t)$ se obtiene al tomar las transformadas de Fourier en ambos miembros de la ecuación (2.48). De ese modo tenemos

$$S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta)[\delta(f - f_c - nf_m) + \delta(f + f_c + nf_m)] \tag{2.49}$$

En la figura 2.23 hemos graficado la función de Bessel $J_n(\beta)$ en función del índice de modulación β para diferentes valores enteros positivos de n. Es posible investigar aún más el

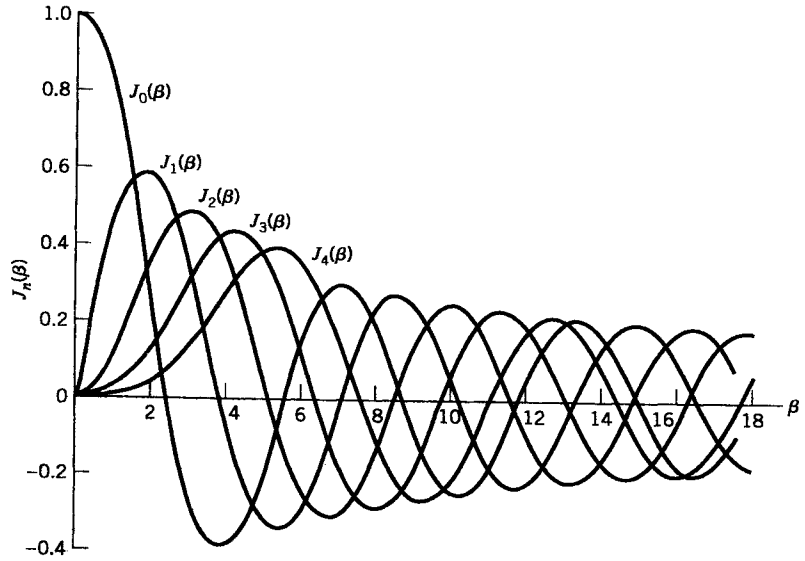


FIGURA 2.23 Gráficas de funciones de Bessel del primer tipo para orden variable.

comportamiento de la función de Bessel $J_n(\beta)$ utilizando las siguientes propiedades (vea el apéndice 3 para mayores detalles):

1. $J_n(\beta) = (-1)^n J_{-n}(\beta)$ para toda n, tanto positiva como negativa (2.50)
2. Para valores pequeños del índice de modulación β, tenemos

$$\left.\begin{aligned} J_0(\beta) &\simeq 1 \\ J_1(\beta) &\simeq \frac{\beta}{2} \\ J_n(\beta) &\simeq 0, \quad n > 2 \end{aligned}\right\} \tag{2.51}$$

3. $$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1 \tag{2.52}$$

Por tanto, al recurrir a las ecuaciones (2.49)-(2.52) y a las curvas de la figura 2.23, podemos hacer las siguientes observaciones:

1. El espectro de una señal de FM contiene una componente de la portadora y un conjunto infinito de frecuencias laterales localizadas simétricamente a cualquier lado de la portadora y a separaciones de frecuencia de f_m, $2f_m$, $3f_m$, $\cdots$. A este respecto, el resultado es distinto al que prevalece en un sistema de AM, ya que en éste una señal moduladora senoidal originaría únicamente un par de frecuencias laterales.
2. En el caso especial de β pequeña comparada con la unidad, sólo los coeficientes de Bessel $J_0(\beta)$ y $J_1(\beta)$ tienen valores significativos, de manera que la señal de FM está efectivamente compuesta por una portadora y un par único de frecuencias laterales en $f_c \pm f_m$. Esta situación corresponde al caso especial de FM de banda angosta que se consideró antes.
3. La amplitud de la componente de la portadora varía con β de acuerdo con $J_0(\beta)$. Esto es, a diferencia de la señal de AM, la amplitud de la componente de la portadora de una señal de FM depende del índice de modulación β. La explicación física de esta propiedad es que la envolvente de una señal de FM es constante, por lo que la potencia promedio de una señal de este tipo que se desarrolla a través de un resistor de 1 ohm también es constante, según indica

$$P = \frac{1}{2} A_c^2 \tag{2.53}$$

Cuando se modula la portadora para generar la señal de FM, la potencia en las frecuencias laterales quizás aparezca sólo a costa de la potencia originalmente en la portadora, lo que ocasiona que la amplitud de la componente de la portadora dependa de β. Advierta que la potencia promedio de una señal de FM también puede determinarse a partir de la ecuación (2.48), obteniéndose

$$P = \frac{1}{2} A_c^2 \sum_{n=-\infty}^{\infty} J_n^2(\beta) \tag{2.54}$$

Al sustituir la ecuación (2.52) en la (2.54), la expresión para la potencia promedio P se reduce a la ecuación (2.53), como debe ser.

▶ Ejemplo 2.2

En este ejemplo deseamos investigar las maneras en las que las variaciones en la amplitud y en la frecuencia de una señal moduladora senoidal afectan el espectro de la señal de FM. Considere

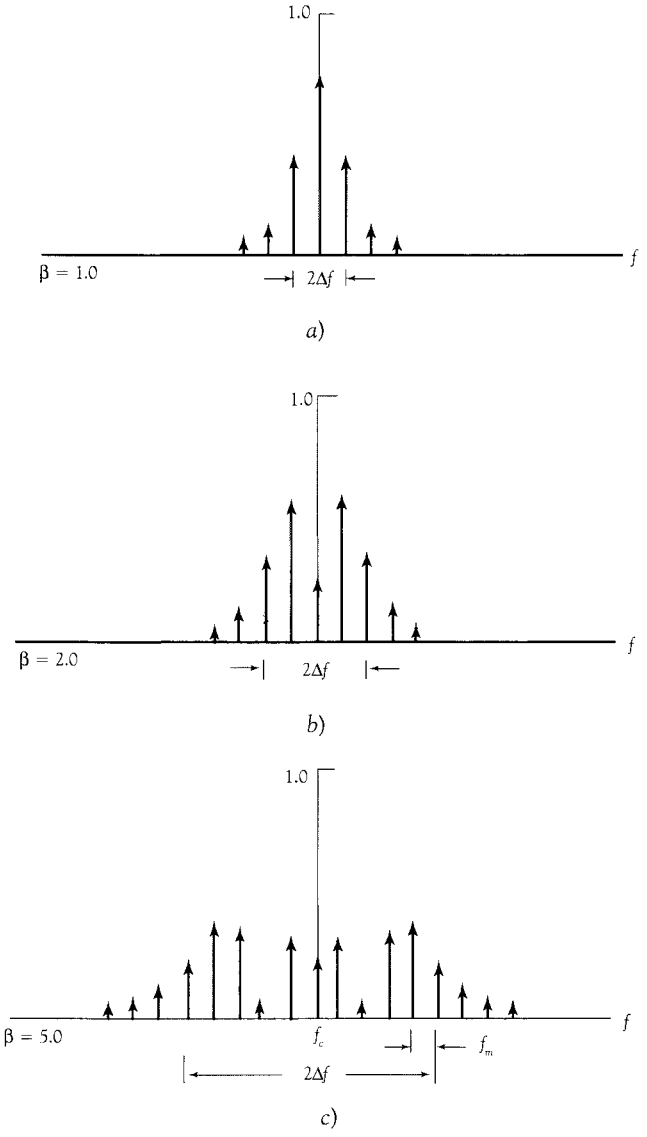


FIGURA 2.24 Espectros de amplitud discretos de una señal de FM normalizada con respecto a la amplitud de la portadora, para el caso de modulación senoidal de frecuencia fija y amplitud variable. Sólo se muestran los espectros para frecuencias positivas.

primero el caso en el que es fija la frecuencia de la señal moduladora, pero que cambia su amplitud, produciéndose una variación correspondiente en la desviación de frecuencia Δf. De tal manera, manteniendo fija la frecuencia de modulación f_m, encontramos que el espectro de amplitud de la señal de FM que resulta es como se muestra graficado en la figura 2.24 para $\beta = 1$, 2 y 5. En este diagrama hemos normalizado el espectro con respecto a la amplitud de portadora no modulada.

Considere a continuación el caso cuando la amplitud de la señal modulada es fija; esto es, se mantiene constante la desviación de frecuencia Δf y se varía la frecuencia de modulación f_m. Determinamos en este caso que el espectro de amplitud de la señal de FM que se produce es como la forma que se grafica en la figura 2.25 para $\beta = 1$, 2 y 5. Vemos que cuando Δf se fija y β aumenta, tenemos un número creciente de líneas espectrales que se apiñan en el intervalo de frecuencia fijo

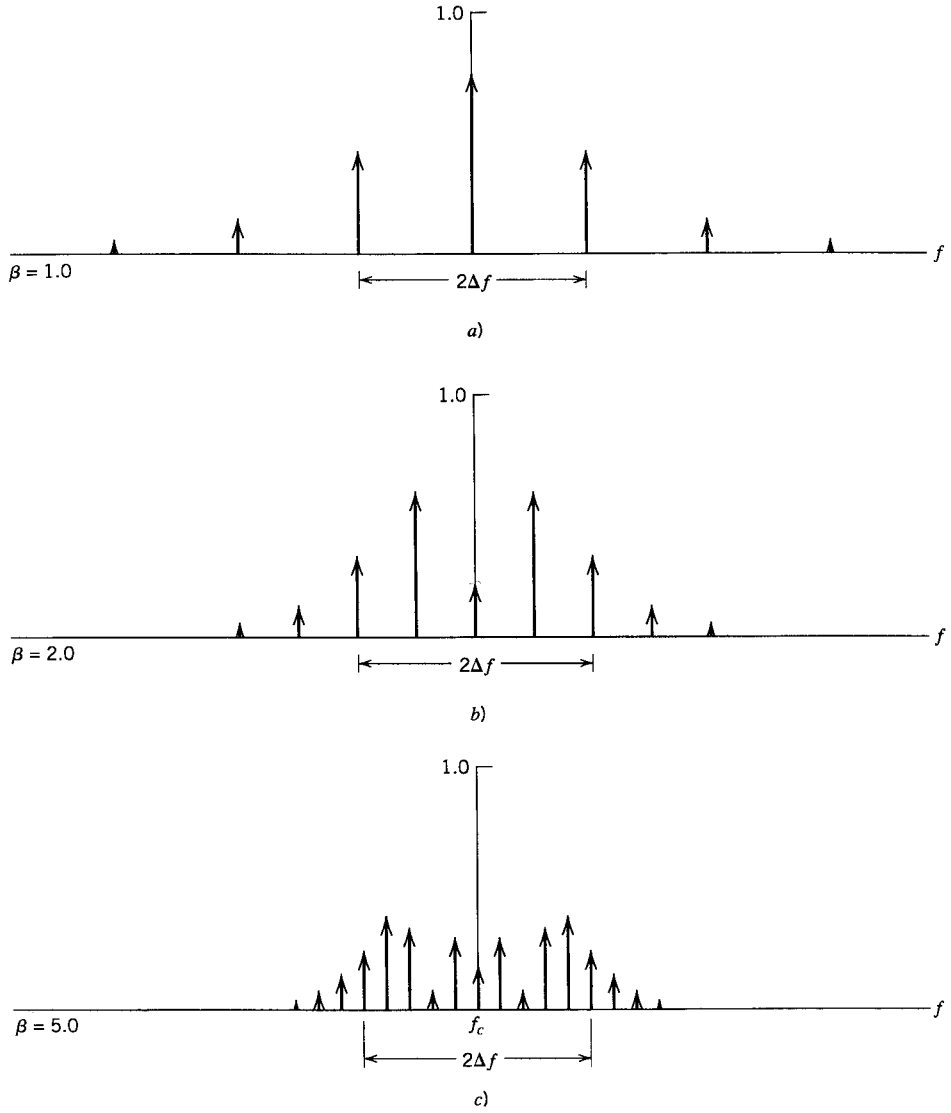


FIGURA 2.25 Espectros de amplitud discretos de una señal de FM normalizada con respecto a la amplitud de la portadora, para el caso de modulación senoidal de frecuencia variable y amplitud fija. Sólo se muestran los espectros de frecuencias positivas.

$f_c - \Delta f < |f| < f_c + \Delta f$. Esto es, cuando β tiende a infinito, el ancho de banda de la onda de FM se acerca al valor límite de $2\Delta f$, que es un punto importante para recordar más adelante. ◀

ANCHO DE BANDA DE TRANSMISIÓN DE SEÑALES DE FM

En teoría, una señal de FM contiene un número infinito de frecuencias laterales de modo que el ancho de banda que se requiere para transmitir este tipo de señales es similarmente infinito en extensión. No obstante, en la práctica encontramos que la señal de FM está efectivamente limitada a un número finito de frecuencias laterales significativas compatible con la cantidad de distorsión especificada. Por tanto es posible que especifiquemos un ancho de banda efectivo requerido para la transmisión de una señal de FM. Considere primero el caso de una señal de

FM generada por una onda moduladora de un tono de frecuencia f_m. En una señal de FM de estas características, las frecuencias laterales que están separadas de la frecuencia de la portadora f_c por una cantidad mayor que la desviación de frecuencia Δf disminuyen rápidamente a cero, por lo que el ancho de banda siempre supera a la excursión total de frecuencia, pero que de cualquier modo está limitado. Específicamente, para valores grandes del índice de modulación β, el ancho de banda tiende, y es sólo un poco mayor que, a la excursión total de frecuencia $2\Delta f$ de acuerdo con la situación que se indica en la figura 2.25. Por otra parte, en el caso de valores pequeños del índice de modulación β, el espectro de la señal de FM está efectivamente limitado a la frecuencia de la portadora f_c y a un par de frecuencias laterales en $f_c \pm f_m$, de modo que el ancho de banda tiende a $2f_m$. Consecuentemente, es posible que definamos una regla aproximada para el ancho de banda de transmisión de una señal de FM generada por una señal moduladora de un tono de frecuencia f_m de la forma siguiente:

$$B_T \simeq 2\Delta f + 2f_m = 2\Delta f\left(1+\frac{1}{\beta}\right) \tag{2.55}$$

Esta relación empírica se conoce como la *regla de Carson.*[4]

Para una evaluación alternativa del requerimiento de ancho de banda de una señal de FM, es posible emplear una definición basada en la retención del número máximo de frecuencias laterales significativas cuyas amplitudes son en su totalidad mayores que algún valor elegido. Una elección conveniente para este valor es 1% de la amplitud de la portadora no modulada. De tal modo, podemos definir el *ancho de banda de transmisión de una onda de FM como la separación entre las dos frecuencias más allá de las cuales ninguna de las frecuencias laterales es mayor que 1% de la amplitud de portadora que se obtiene cuando se elimina la modulación.* Esto es, definimos el ancho de banda de transmisión como $2n_{máx}f_m$, donde f_m es la frecuencia de modulación y $n_{máx}$ es el valor más grande del entero n que satisface el requerimiento $|J_n(\beta)| > 0.01$. El valor de $n_{máx}$ varía con el índice de modulación β y puede determinarse fácilmente de valores tabulados de la función de Bessel $J_n(\beta)$. La tabla 2.2 presenta el número total de frecuencias laterales significativas (incluyendo ambas frecuencias laterales superior e inferior) para diferentes valores de β, calculadas sobre la base de 1% que se explicó aquí. El ancho de banda de transmisión B_T que se calculó al utilizar este procedimiento puede presentarse en la forma de una *curva universal* normalizándola respecto de la desviación de frecuencia Δf y graficándola después en función de β. Esta curva se presenta en la figura 2.26, la cual se dibuja como el mejor ajuste a través del conjunto de puntos obtenidos al utilizar la tabla 2.2. En la figura 2.26 advertimos que cuando aumenta el índice de modulación β, el ancho de banda que ocupan las frecuencias

TABLA 2.2 ***Número de frecuencias laterales significativas de una señal de FM de banda ancha para índice de modulación variable***

Índice de modulación β	*Número de frecuencias laterales significativas* $2n_{máx}$
0.1	2
0.3	4
0.5	4
1.0	6
2.0	8
5.0	16
10.0	28
20.0	50
30.0	70

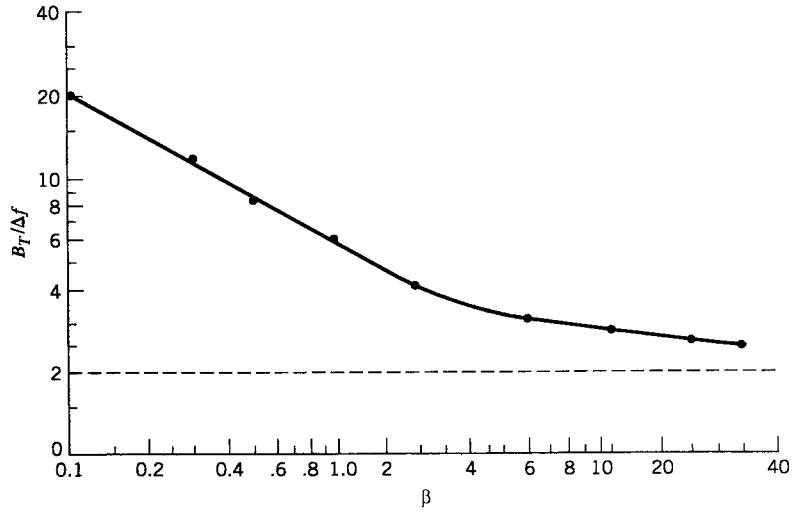


FIGURA 2.26 Curva universal para evaluar el ancho de banda de 1% de una onda de FM.

laterales significativas disminuye hacia aquel a partir del cual se desvía en realidad la frecuencia de la portadora. Esto quiere decir que valores pequeños del índice de modulación β resultan relativamente más extraños en el ancho de banda de transmisión que los valores mayores de β.

Considere a continuación el caso más general de una señal moduladora arbitraria $m(t)$ con su componente de frecuencia más alta denotada por W. El ancho de banda que se requiere para transmitir una señal de FM generada por esta señal moduladora se estima utilizando un análisis de modulación de tono correspondiente al peor caso. De modo específico, establecemos primero la llamada *razón de desviación* D, que se define como la proporción entre la desviación de la frecuencia Δf, que corresponde a la máxima amplitud posible de la señal de modulación $m(t)$, y la frecuencia de modulación más alta W. Estas condiciones representan los casos extremos posibles. *La razón de desviación D desempeña el mismo papel en la modulación no senoidal que el del índice de modulación* β *en el caso de modulación senoidal.* Por tanto, al sustituir β por D y f_m por W, es posible utilizar la regla de Carson dada por la ecuación (2.55) o la curva universal de la figura 2.26 para obtener un valor correspondiente al ancho de banda de transmisión de la señal de FM. Desde un punto de vista práctico, la regla de Carson subestima un poco el requerimiento de ancho de banda de un sistema de FM, en tanto que al usar la curva universal de la figura 2.26 se obtiene un resultado algo conservador. Por tanto, la elección de un ancho de banda de transmisión que se encuentra entre los límites que ofrecen estas dos reglas prácticas resulta aceptable para la mayoría de los fines prácticos.

EJEMPLO 2.3

En Estados Unidos de Norteamérica, el valor máximo de la desviación de frecuencia Δf se fija en 75 kHz para la transmisión comercial de radio FM. Si consideramos la frecuencia de modulación W = 15 kHz, que es por lo común la frecuencia de audio "máxima" de interés en las transmisiones de FM, encontramos que el valor correspondiente de la razón de desviación es

$$D = \frac{75}{15} = 5$$

Al utilizar la regla de Carson de la ecuación (2.55), remplazando β por D y f_m por W, el valor aproximado del ancho de banda de transmisión de la señal de FM se obtiene como

$$B_T = 2(75 + 15) = 180 \text{ kHz}$$

Por otro lado, el uso de la curva de la figura 2.26 produce un ancho de banda de transmisión de la señal de FM igual a

$$B_T = 3.2\Delta f = 3.2 \times 75 = 240 \text{ kHz}$$

En la práctica se asigna un ancho de banda de 200 kHz a cada transmisor de FM. Con base en esto, la regla de Carson subestima el ancho de banda de transmisión en 10%, en tanto que la curva universal de la figura 2.26 la sobreestima en 20%. ◀

■ GENERACIÓN DE SEÑALES DE FM

Existen en esencia dos métodos básicos de generación de señales moduladas en frecuencia, a saber, FM *directa* y FM *indirecta*. En el método directo la frecuencia de la portadora se varía directamente de acuerdo con la señal de banda base de entrada, lo cual se consigue fácilmente utilizando un *oscilador controlado por voltaje*. En el método indirecto la señal moduladora se emplea primero para producir una señal de FM de banda angosta y a continuación se utiliza la multiplicación de frecuencia para aumentar la desviación de frecuencia en el nivel deseado. El método indirecto es el que se prefiere para la modulación en frecuencia cuando la estabilidad de la frecuencia de la portadora es la principal preocupación, como en el caso de la transmisión de radio comercial, según se describe a continuación.

FM indirecta[5]

En la figura 2.27 se presenta un diagrama de bloques simplificado y un sistema de FM indirecta. La señal $m(t)$ del mensaje (banda de base) se integra primero y luego se utiliza para modular en fase un oscilador controlado por cristal; el uso de este último proporciona *estabilidad de frecuencia*. Para minimizar la distorsión inherente en el modulador de fase, la desviación de fase máxima o el índice de modulación β se mantiene pequeño produciéndose de ese modo una señal de FM de banda angosta; para la puesta en práctica del modulador de fase de banda angosta, podemos utilizar el arreglo que se describe en la figura 2.21. La señal de FM de banda angosta se multiplica después en frecuencia por medio de un multiplicador de frecuencia hasta que se produce la señal de FM de banda ancha que se desea.

Un *multiplicador de frecuencia* está compuesto por un dispositivo no lineal seguido por un filtro pasobanda, como se indica en la figura 2.28. La implicación de que el dispositivo no lineal sea sin memoria corresponde a que no cuenta con elementos que almacenan energía. La relación de entrada-salida de un dispositivo de este tipo puede expresarse en la forma general

$$v(t) = a_1 s(t) + a_2 s^2(t) + \ldots + a_n s^n(t) \tag{2.56}$$

donde $a_1, a_2, \ldots, a_n$ son coeficientes determinados por el punto de operación del dispositivo y n es la *no linealidad de mayor orden*. En otras palabras, el dispositivo no lineal sin memoria es uno de potencia n-ésima. La entrada $s(t)$ es una señal de FM definida por

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)d\tau\right]$$

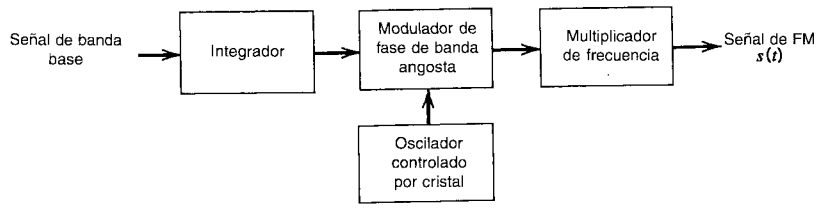


FIGURA 2.27 Diagrama de bloques del método indirecto de generar una señal de FM de banda ancha.

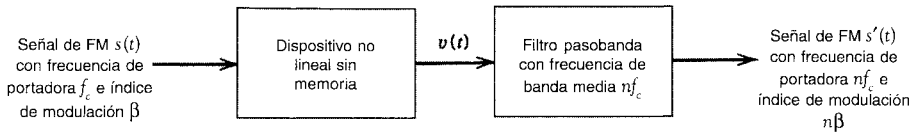


FIGURA 2.28 Diagrama de bloques de un multiplicador de frecuencia.

cuya frecuencia instantánea es

$$f_i(t) = f_c + k_f m(t) \tag{2.57}$$

La frecuencia de banda media del filtro pasobanda de la figura 2.28 se iguala a nf_c, donde f_c es la frecuencia de la portadora de la señal de FM entrante $s(t)$. Además, el filtro pasobanda se diseña para que tenga un ancho de banda igual a n veces el ancho de banda de transmisión de $s(t)$. En la sección 2.8 que aborda los efectos no lineales en los sistemas de FM, escribimos las contribuciones espectrales de tales términos no lineales como los de segundo y tercer orden en la relación de entrada-salida de la ecuación (2.56). Por ahora es suficiente afirmar que después del filtrado pasobanda de la salida del dispositivo no lineal $v(t)$, tenemos una nueva señal de FM definida por

$$s'(t) = A_c' \cos\left[2\pi n f_c t + 2\pi n k_f \int_0^t m(\tau)\, d\tau\right] \tag{2.58}$$

cuya frecuencia instantánea es

$$f_i'(t) = nf_c + nk_f m(t) \tag{2.59}$$

De tal modo, comparando la ecuación (2.59) con la (2.57), vemos que el circuito de procesamiento no lineal de la figura 2.28 actúa como un multiplicador de frecuencia. La razón de multiplicación de frecuencia se determina a partir de la potencia n más elevada en la relación de entrada-salida de la ecuación (2.56), lo que caracteriza al dispositivo no lineal sin memoria.

■ DEMODULACIÓN DE SEÑALES DE FM

La *demodulación de frecuencia* es el proceso que nos permite recuperar la señal moduladora original a partir de una señal modulada de frecuencia. El objetivo es producir una característica de transferencia que sea la inversa de la del modulador de frecuencia, la cual puede realizarse directa o indirectamente. Aquí describimos un método directo de demodulación en frecuencia que incluye el uso de un popular dispositivo conocido como discriminador de frecuencia, cuya amplitud de salida instantánea es directamente proporcional a la frecuencia instantánea de la señal de FM de entrada. En la sección 2.14 describimos un método indirecto de demodulación de frecuencia que utiliza otro dispositivo popular que recibe el nombre de malla de fase encadenada.

Básicamente, el *discriminador de frecuencia* consiste en un *circuito de pendiente* seguido por un *detector de envolvente*. Un circuito de pendiente ideal se caracteriza por una respuesta en frecuencia que es puramente imaginaria, variando linealmente con la frecuencia dentro de un intervalo preestablecido de esta misma. Considere la respuesta en frecuencia que se indica en la figura 2.29*a*, la cual está definida por

$$H_1(f) = \begin{cases} j2\pi a\left(f - f_c + \dfrac{B_T}{2}\right), & f_c - \dfrac{B_T}{2} \le f \le f_c + \dfrac{B_T}{2} \\ j2\pi a\left(f + f_c - \dfrac{B_T}{2}\right), & -f_c - \dfrac{B_T}{2} \le f \le -f_c + \dfrac{B_T}{2} \\ 0, & \text{en otro lado} \end{cases} \tag{2.60}$$

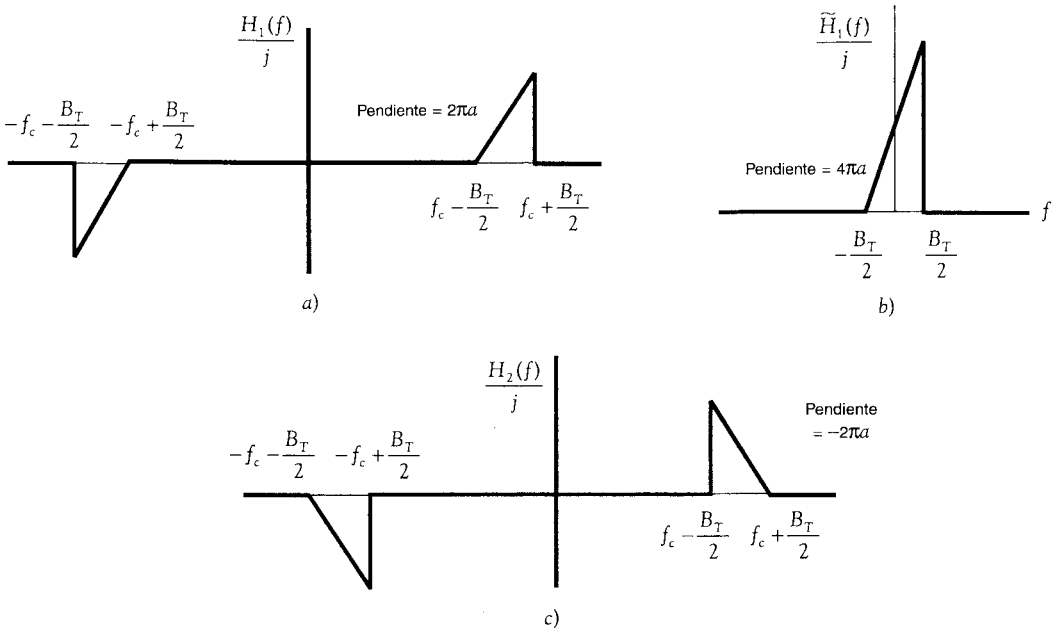


FIGURA 2.29 *a*) Respuesta en frecuencia de un circuito de pendiente ideal. *b*) Respuesta en frecuencia del equivalente complejo pasobajas del circuito de pendiente. *c*) Respuesta en frecuencia del circuito de pendiente complementario al de la parte *a*.

donde a es una constante. Nos interesa evaluar la respuesta de este circuito de pendiente denotada por $s_1(t)$, la cual es producida por una señal de FM $s(t)$ de frecuencia de portadora f_c y ancho de banda de transmisión B_T. Se supone que el espectro de $s(t)$ es esencialmente cero fuera del intervalo de frecuencia $f_c - B_T/2 \leq |f| \leq f_c + B_T/2$.

Podemos simplificar el análisis del discriminador de frecuencia apelando al isomorfismo entre el filtro pasobanda de valores reales y un filtro pasobajas de valores complejos. Este isomorfismo se explica en el apéndice 2. De acuerdo con el material que se presenta en ese apéndice, resulta factible reemplazar el filtro pasobanda con respuesta en frecuencia $H_1(f)$ por un filtro pasobajas equivalente con respuesta en frecuencia $\widetilde{H}_1(f)$ efectuando dos cosas:

1. Desplazamos $\widetilde{H}_1(f)$ a la derecha en f_c, donde f_c es la frecuencia de banda media del filtro pasobanda; esta operación alinea la respuesta en frecuencia trasladada del filtro pasobajas equivalente con la del filtro pasobanda.
2. Igualamos $\widetilde{H}_1(f - f_c)$ a $2H_1(f)$ para $f > 0$.

De modo que para el problema presente obtenemos

$$\widetilde{H}_1(f - f_c) = 2H_1(f), \quad f > 0 \tag{2.61}$$

En consecuencia, al utilizar las ecuaciones (2.60) y (2.61), encontramos

$$\widetilde{H}_1(f) = \begin{cases} j4\pi a\left(f + \dfrac{B_T}{2}\right), & -\dfrac{B_T}{2} \leq f \leq \dfrac{B_T}{2} \\ 0, & \text{en otro lado} \end{cases} \tag{2.62}$$

la cual se grafica en la figura 2.29*b*.

La señal de FM entrante $s(t)$ está definida por la ecuación (2.26), misma que se reproduce aquí por conveniencia:

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\, d\tau\right]$$

Debido a que la frecuencia de portadora f_c es alta en comparación con el ancho de banda de transmisión de la señal de FM $s(t)$, la envolvente compleja de $s(t)$ es

$$\tilde{s}(t) = A_c \exp\left[j2\pi k_f \int_0^t m(\tau)\, d\tau\right] \tag{2.63}$$

Dejemos que $\tilde{s}_1(t)$ denote la envolvente compleja de la respuesta del circuito de pendiente definido por la figura 2.29*a* y que se debe a $\tilde{s}(t)$. Por tanto, de acuerdo con el material que se presenta en el apéndice 2, podemos expresar la transformada de Fourier de $\tilde{s}_1(t)$ del modo siguiente:

$$\begin{aligned}\tilde{S}_1(f) &= \frac{1}{2}\tilde{H}_1(f)\tilde{S}(f) \\ &= \begin{cases} j2\pi a\left(f + \dfrac{B_T}{2}\right)\tilde{S}(f), & -\dfrac{B_T}{2} \le f \le \dfrac{B_T}{2} \\ 0, & \text{en otro lado} \end{cases}\end{aligned} \tag{2.64}$$

donde $\tilde{S}(f)$ es la transformada de Fourier de $\tilde{s}(t)$. De acuerdo con el análisis de Fourier, sabemos que la multiplicación de la transformada de Fourier de una señal por $j2\pi f$ es equivalente a diferenciarla en el dominio del tiempo; vea la línea 8 de la tabla A6.2. Por tanto, de acuerdo con la ecuación (2.64), deducimos

$$\tilde{s}_1(t) = a\left[\frac{d\tilde{s}(t)}{dt} + j\pi B_T \tilde{s}(t)\right] \tag{2.65}$$

Al sustituir la ecuación (2.63) en la (2.65), obtenemos

$$\tilde{s}_1(t) = j\pi B_T a A_c\left[1 + \frac{2k_f}{B_T} m(t)\right] \exp\left[j2\pi k_f \int_0^t m(\tau)\, d\tau\right] \tag{2.66}$$

La respuesta deseada del circuito de pendiente es entonces

$$\begin{aligned} s_1(t) &= \mathrm{Re}\left[\tilde{s}_1(t) \exp(j2\pi f_c t)\right] \\ &= \pi B_T a A_c\left[1 + \frac{2k_f}{B_T} m(t)\right] \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\, d\tau + \frac{\pi}{2}\right] \end{aligned} \tag{2.67}$$

La señal $s_1(t)$ es una señal modulada híbrida en la que tanto la amplitud como la frecuencia de la onda portadora varían con la señal del mensaje $m(t)$. Sin embargo, en vista de que elegimos

$$\left|\frac{2k_f}{B_T} m(t)\right| < 1 \quad \text{para todo } t$$

es viable que usemos en ese caso un detector de envolvente para recuperar las variaciones de amplitud y obtener así, salvo por un término de polarización, la señal del mensaje original. En consecuencia, la salida del detector de envolvente resultante es

$$\left|\tilde{s}_1(t)\right| = \pi B_T a A_c\left[1 + \frac{2k_f}{B_T} m(t)\right] \tag{2.68}$$

El término de polarización $\pi B_T a A_c$ en el segundo miembro de la ecuación (2.68) es proporcional a la pendiente a de la función de transferencia del circuito de pendiente. Esto sugiere que

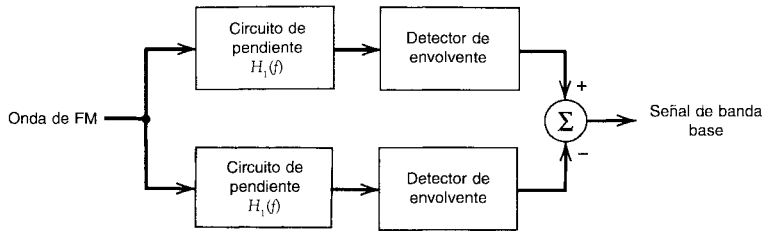


FIGURA 2.30 Diagrama de bloques de discriminador de frecuencia.

la polarización quizá se elimine al sustraer de la salida del detector de envolvente $|\tilde{s}_1(t)|$ la salida de un segundo detector de envolvente precedida por el *circuito de pendiente complementario*, con una respuesta en frecuencia $H_2(f)$ como se describe en la figura 2.29c. Esto es, los dos circuitos de pendiente se relacionan por medio de

$$\tilde{H}_2(f) = \tilde{H}_1(-f) \tag{2.69}$$

Dejemos que $s_2(t)$ denote la respuesta del circuito de pendiente complementario que produce la señal de FM entrante $s(t)$. En estas condiciones, siguiendo el procedimiento similar al que acaba de describirse, es posible escribir

$$|\tilde{s}_2(t)| = \pi B_T a A_c \left[1 - \frac{2k_f}{B_T} m(t)\right] \tag{2.70}$$

donde $s_2(t)$ es la envolvente compleja de la señal $s_2(t)$. La diferencia entre las dos envolventes en la ecuación (2.68) y (2.70) es

$$\begin{aligned} s_o(t) &= |\tilde{s}_1(t)| - |\tilde{s}_2(t)| \\ &= 4\pi k_f a A_c m(t) \end{aligned} \tag{2.71}$$

que es la versión escalada de la señal del mensaje original $m(t)$ y sin polarización.

En consecuencia, es posible modelar el *discriminador de frecuencia ideal* como un par de circuitos de pendiente con sus funciones de transferencia compleja relacionadas por la ecuación (2.69), seguido por detectores de envolvente y al final por un sumador, como en la figura 2.30. Este esquema recibe el nombre de *discriminador de frecuencia balanceado.*

■ MULTIPLEXADO ESTEREOFÓNICO DE FM[6]

El *multiplexado estereofónico* es una forma de multiplexado por división de frecuencia (FDM) diseñado para transmitir dos señales independientes por la misma portadora. Se usa ampliamente en la transmisión de radio de FM para enviar dos elementos diferentes de un programa (por ejemplo, dos secciones diferentes de una orquesta, un vocalista y un acompañante), de manera que proporcione una dimensión espacial para la percepción de un oyente en el extremo de recepción.

La especificación de los estándares para la transmisión estéreo de FM recibe la influencia de dos factores:

1. La transmisión tiene que operar dentro de los canales asignados de transmisión de FM.
2. Tiene que ser compatible con los receptores de radio monofónicos.

El primer requerimiento establece los parámetros de frecuencia permisibles, que incluyen la desviación de frecuencia. El segundo requerimiento restringe la manera en la cual se configura la señal transmitida.

La figura 2.31*a* presenta el diagrama de bloques del sistema de multiplexado que se utiliza en un transmisor estereofónico de FM. Sean $m_l(t)$ y $m_r(t)$ las señales captadas por los micrófonos

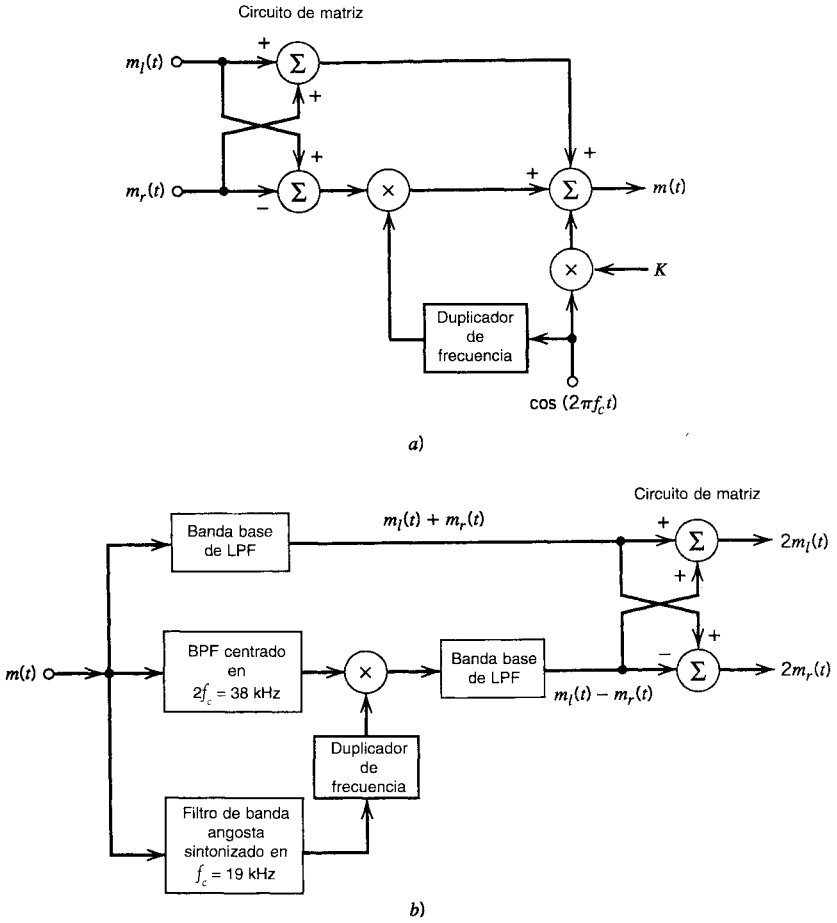


FIGURA 2.31 *a*) Multiplexor en el FM estereofónico. *b*) Demultiplexor de FM estereofónico.

izquierdo y derecho en el extremo de transmisión del sistema. Éstas se aplican a un *circuito de matriz* simple que genera la *señal de suma*, $m_l(t) + m_r(t)$, y la *señal de diferencia*, $m_l(t) - m_r(t)$. La señal de suma no se procesa y se deja en su forma de banda base; ésta se encuentra disponible para la recepción monofónica. La señal de diferencia y la subportadora de 38 kHz (obtenida de un oscilador a cristal de 19 kHz por medio de duplicación de frecuencia) se aplican a un modulador de producto, produciéndose de ese modo una onda modulada de DSB-SC. Además de la señal de suma y de esta onda modulada de DSB-SC, la señal multiplexada $m(t)$ incluye también un piloto de 19 kHz para ofrecer una referencia en la detección coherente de la señal de diferencia en el receptor estéreo. De esa manera, la señal multiplexada se describe por medio de

$$m(t) = [m_l(t) + m_r(t)] + [m_l - m_r(t)]\cos(4\pi f_c t) + K\cos(2\pi f_c t) \tag{2.72}$$

donde $f_c = 19$ kHz y K es la amplitud del tono piloto. En esas condiciones, la señal multiplexada $m(t)$ modula en frecuencia la portadora principal para producir la señal transmitida. El piloto se distribuye entre 8 y 10 por ciento de la desviación de frecuencia pico; la amplitud K en la ecuación (2.72) se elige para satisfacer este requerimiento.

En un receptor estereofónico, la señal multiplexada $m(t)$ se recupera demodulando en frecuencia la onda de FM entrante. Después $m(t)$ se aplica en el *sistema de demultiplexado* que se

muestra en la figura 2.31*b*. Las componentes individuales de la señal multiplexada $m(t)$ se separan utilizando tres filtros apropiados. El piloto recuperado (utilizando un filtro de banda angosta sintonizado en 19 kHz) se duplica en frecuencia para producir la subportadora deseada de 38 kHz. La disponibilidad de esta subportadora permite la detección coherente de la onda modulada de DSB-SC, recuperándose de esa manera la señal de diferencia, $m_l(t) - m_r(t)$. El filtro pasobanda de la banda base en la trayectoria superior de la figura 2.31*b* se diseña para dejar pasar la señal de suma $m_l(t) + m_r(t)$. Por último, el circuito de matriz simple reconstruye la señal de lado izquierdo $m_l(t)$ y la señal de lado derecho $m_r(t)$, excepto por factores de escalamiento, y éstas se aplican a sus respectivos altavoces.

2.8 *Efectos no lineales en sistemas de FM*

En las dos secciones anteriores estudiamos la teoría de modulación de frecuencia y los métodos para su generación y demodulación. Completamos la exposición de este tema considerando los efectos no lineales en los sistemas de FM.

Las *no linealidades* están presentes, de una u otra forma, en todas las redes eléctricas. Existen dos formas básicas de no linealidad que considerar:

1. Se afirma que la no linealidad es *fuerte* cuando se introduce intencionalmente y de una manera controlada en alguna aplicación específica. Entre los ejemplos de la no linealidad fuerte se incluyen los moduladores cuadráticos, los limitadores rígidos y los multiplicadores de frecuencia.
2. Se dice que la no linealidad es *débil* cuando se desea un desempeño lineal, pero surgen no linealidades de naturaleza parásita debido a imperfecciones. El efecto de estas no linealidades débiles es limitar los niveles de señal útiles en un sistema y, en consecuencia, se vuelven una importante consideración del diseño.

En esta sección examinamos el efecto de no linealidades débiles sobre la modulación de frecuencia.

Considere un canal de comunicaciones, cuya característica de transferencia se define por medio de la relación de entrada-salida no lineal

$$v_o(t) = a_1 v_i(t) + a_2 v_i^2(t) + a_3 v_i^3(t) \tag{2.73}$$

donde $v_i(t)$ y $v_0(t)$ son las señales de entrada y salida, respectivamente, y a_1, a_2 y a_3 son constantes; la ecuación (2.73) es una versión truncada de la ecuación (2.56) utilizada en el contexto de la multiplicación de frecuencia. El canal que se describe en la ecuación (2.73) se dice que es *sin memoria* en cuanto a que la señal de salida $v_o(t)$ es una función instantánea de la señal de entrada $v_i(t)$ (es decir, no hay almacenamiento de energía implicado en la descripción). Deseamos determinar el efecto de transmitir una onda modulada en frecuencia por un canal de este tipo. La señal de FM se define mediante

$$v_i(t) = A_c \cos[2\pi f_c t + \phi(t)]$$

donde

$$\phi(t) = 2\pi k_f \int_0^t m(\tau) d\tau$$

Para esta señal de entrada, el uso de la ecuación (2.73) produce

$$\begin{aligned} v_o(t) = {} & a_1 A_c \cos[2\pi f_c t + \phi(t)] + a_2 A_c^2 \cos^2[2\pi f_c t + \phi(t)] \\ & + a_3 A_c^3 \cos^3[2\pi f_c t + \phi(t)] \end{aligned} \tag{2.74}$$

Expandiendo los términos coseno al cuadrado y al cubo en la ecuación (2.74) y agrupando términos comunes, obtenemos

$$\begin{aligned} v_o(t) = &\frac{1}{2}a_2A_c^2 + \left(a_1A_c + \frac{3}{4}a_3A_c^3\right)\cos[2\pi f_c t + \phi(t)] \\ &+ \frac{1}{2}a_2A_c^2\cos[4\pi f_c t + 2\phi(t)] \\ &+ \frac{1}{4}a_3A_c^3\cos[6\pi f_c t + 3\phi(t)] \end{aligned} \tag{2.75}$$

De tal modo, la salida del canal consiste en una componente de cd y en tres señales moduladas en frecuencia con frecuencias de portadora f_c, $2f_c$ y $3f_c$; los últimos componentes se establecen mediante la aportación respectiva de los términos lineales de segundo y tercer orden de la ecuación (2.73).

Para extraer la señal de FM deseada de la salida del canal $v_o(t)$, es decir, la componente particular con frecuencia de portadora f_c, es necesario separar la señal de FM con esta frecuencia de portadora de aquella con la frecuencia de portadora más cercana, $2f_c$. Dejemos que Δf denote la desviación de frecuencia de la señal de FM entrante $v_i(t)$ y que W indique la componente de más alta frecuencia de la señal del mensaje $m(t)$. En ese caso, al aplicar la regla de Carson y al observar que se duplica la desviación de frecuencia alrededor de la segunda armónica de la frecuencia de portadora, encontramos que la condición necesaria para separar la señal de FM deseada con la frecuencia de portadora f_c de aquella con frecuencia de portadora $2f_c$ es

$$2f_c - (2\Delta f + W) > f_c + \Delta f + W$$

o bien,

$$f_c > 3\Delta f + 2W \tag{2.76}$$

Por tanto, mediante el uso de un filtro pasobanda de frecuencia de banda media f_c y ancho de banda $2\Delta f + 2W$, la salida del canal se reduce en

$$v_o'(t) = \left(a_1A_c + \frac{3}{4}a_3A_c^3\right)\cos[2\pi f_c t + \phi(t)] \tag{2.77}$$

Vemos, por tanto, que el único efecto de dejar pasar una señal de FM por un canal con no linealidades de amplitud, seguida por el filtrado apropiado, implica simplemente modificar su amplitud. Esto es, a diferencia de la modulación de amplitud, la modulación de frecuencia no es afectada por la distorsión producida en la transmisión por un canal con no linealidades de amplitud. Ésta es la razón por la que encontramos la aplicación de la modulación de frecuencia en los sistemas de radio de microondas: posibilita el uso de amplificadores y transmisores de potencia altamente no lineales, mismos que son en particular importantes para producir una salida de potencia máxima a las frecuencias de radio.

Sin embargo, un sistema de FM es en extremo sensible a las *no linealidades de fase*, como esperaríamos intuitivamente. Un tipo común de no linealidad de fase que se encuentra en los sistemas de radio de microondas se conoce como *conversión de AM a PM*. Éste es el resultado de las características de fase de repetidores o amplificadores utilizados en el sistema que son dependientes de la amplitud instantánea de la señal de entrada. En la práctica, la conversión de AM-PM se caracteriza por una constante K, que se mide en grados por dB y que puede interpretarse como el cambio de fase máximo en la salida para un cambio de 1 dB en la envolvente a la entrada. Cuando una onda de FM se transmite por un enlace de radio de microondas, ésta capta variaciones espurias de amplitud debidas a ruido e interferencia durante el curso de la transmisión, y cuando tal onda de FM se deja pasar por un repetidor con conversión de AM a PM, la salida contendrá

modulación en fase indeseable y la distorsión correspondiente. Por tanto, es importante mantener a un nivel bajo la conversión AM-PM. Por ejemplo, en un buen repetidor de microondas, la constante de conversión K de AM-PM es menor que 2 grados por dB.

2.9 *Receptor superheterodino*[7]

En un sistema de *transmisión*, sin que importe que éste se base en la modulación de amplitud o en la modulación de frecuencia, el receptor no sólo tiene la tarea de demodular la señal modulada entrante, sino que también se requiere que realice algunas otras funciones del sistema:

- *Sintonización de la frecuencia portadora*, cuyo propósito es elegir la señal deseada (es decir, la estación de radio o TV que se desea).
- *Filtrado*, que se requiere para separar la señal deseada de otras señales moduladas que tal vez se captaron a lo largo del camino.
- *Amplificación*, la cual persigue compensar la pérdida de potencia de la señal a la que se incurre en el curso de la transmisión.

El *receptor superheterodino*, o *superhet* como a menudo se le conoce, es un tipo especial de receptor que satisface las tres funciones, sobre todo las primeras dos, de una manera elegante y práctica. De modo específico, este receptor supera la dificultad de tener que construir un filtro sintonizable variable y altamente selectivo. En realidad, todos los receptores de radio y TV que se fabrican en la actualidad son de este tipo.

Básicamente, el receptor consiste en una sección de radio frecuencia (RF), un mezclador y un oscilador local, una sección de frecuencia intermedia (FI), demodulador y amplificador de potencia. Los parámetros de frecuencia característicos de los receptores de radio comerciales de AM y FM se listan en la tabla 2.3. La figura 2.32 presenta el diagrama de bloques de un receptor superheterodino para modulación de amplitud utilizando un detector de envolvente para la demodulación.

La onda entrante modulada de amplitud se capta por medio de una antena receptora y se amplifica en la sección RF que se sintoniza a la frecuencia de la portadora de la onda entrante. La combinación del mezclador y el oscilador local (de frecuencia ajustable) proporciona una función de *heterodinado*, mediante el cual la señal de entrada se convierte en una *frecuencia intermedia* fija y predeterminada, casi siempre inferior a la frecuencia de la portadora recibida. Esta traslación de frecuencia se consigue sin perturbar la relación de las bandas laterales de la portadora; vea la sección 2.4. El resultado del heterodinado es producir una portadora de frecuencia intermedia definida por

$$f_{FI} = f_{LO} - f_{RF} \tag{2.78}$$

donde f_{LO} es la frecuencia del oscilador local y f_{RF} es la frecuencia de la portadora de la señal RF entrante. Nos referimos a f_{FI} como la frecuencia intermedia (FI), ya que la señal no está a la

TABLA 2.3 *Parámetros de frecuencia característicos de los receptores de radio de AM y FM*

	Radio de AM	*Radio de FM*
Intervalo de portadora de RF	0.535–1.605 MHz	88–108 MHz
Frecuencia de banda media de la sección de FI	0.455 MHz	10.7 MHz
Ancho de banda de la FI	10 kHz	200 kHz

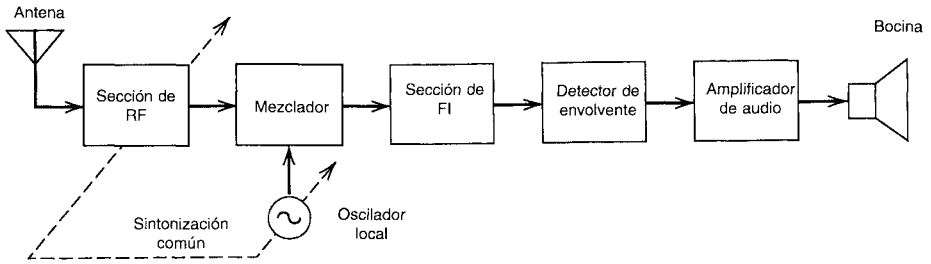


FIGURA 2.32 Elementos básicos de un receptor de radio AM del tipo superheterodino.

frecuencia de entrada original ni a la frecuencia de la banda de base final. La combinación mezclador y oscilador local, algunas veces se conoce como el *detector primero*, en cuyo caso el demodulador se denomina *detector segundo*.

La sección FI está compuesta por una o más etapas de amplificación sintonizada, con un ancho de banda correspondiente al que se requiere para el tipo particular de demodulación que se intenta que maneje el receptor. La sección de FI proporciona la mayor parte de la amplificación y selectividad en el receptor. La salida de la sección FI se aplica a un demodulador, cuyo propósito es recuperar la señal de la banda base. Si se recurre a la detección coherente, entonces debe proporcionarse una fuente de señales coherentes en el receptor. La operación final en este último es la amplificación de potencia de la señal del mensaje recuperada.

En un receptor superheterodino el mezclador creará una salida de frecuencia intermedia cuando la frecuencia de la señal de entrada sea mayor o menor que la frecuencia del oscilador local en una cantidad igual a la frecuencia intermedia. Esto es, existen dos frecuencias de entrada, a saber, $|f_{LO} \pm f_{FI}|$, que producirán f_{FI} en la salida del mezclador. Lo anterior introduce la posibilidad de la recepción simultánea de dos señales que difieren en frecuencia por el doble de la frecuencia intermedia. Por ejemplo, un receptor sintonizado en 0.65 MHz y que cuenta con una FI de 0.455 MHz está sujeto a una *interferencia de imagen* a 1.56 MHz; en realidad, cualquier receptor con este valor de FI, cuando se sintoniza en cualquier estación, se somete a una interferencia de imagen a una frecuencia de 0.910 MHz superior que la de la estación deseada. Puesto que la función del mezclador es producir la diferencia entre dos frecuencias aplicadas, es incapaz de distinguir entre la señal deseada y su imagen en cuanto a que produce una salida de FI a partir de cualquiera de ellas. El único remedio práctico para la interferencia de imagen consiste en emplear etapas altamente selectivas en la sección de RF (es decir, entre la antena y el mezclador) para favorecer la señal deseada y discriminarla con respecto a la *señal de imagen* o indeseada. La eficacia de la supresión de señales de imagen indeseadas aumenta cuando se incrementa el número de etapas selectivas en la sección RF, y cuando crece la proporción entre las frecuencias intermedia y de la señal.

La diferencia básica entre los receptores superheterodinos de AM y FM se encuentra en el uso de un demodulador de FM, tal como el discriminador limitador de frecuencia. En un sistema de FM, las variaciones de la frecuencia instantánea de una onda de portadora senoidal transmiten la información del mensaje, y su amplitud se mantiene constante. Por tanto, cualesquiera variaciones de la amplitud de la portadora en la entrada del receptor deben originarse del ruido o la interferencia. Un *limitador de amplitud*, que sigue a la sección de FI, se utiliza para eliminar las variaciones de amplitud sujetando la onda modulada en la salida de la sección de FI casi en el eje cero. La onda rectangular que resulta se redondea por medio de un filtro pasobanda que suprime las armónicas de la frecuencia de la portadora. De ese modo la salida del filtro es otra vez senoidal, con una amplitud prácticamente independiente de aquella de la portadora en la entrada del receptor (vea el problema 2.42).

2.10 *Ruido en sistemas de modulación de onda continua*

Hasta este punto nuestra exposición se ha centrado en la caracterización de técnicas de modulación de onda continua (CW), desde una perspectiva por completo determinista. En el resto del capítulo estudiamos los efectos del *ruido del canal* en la recepción de señales moduladas de onda continua y desarrollamos de ese modo una amplia comprensión del comportamiento de las comunicaciones analógicas.

Para emprender un estudio de este tipo seguimos la práctica de costumbre al formular dos modelos:

1. *Modelo del canal*, que supone un canal de comunicación que no está distorsionado, pero al cual perturba el *ruido blanco gaussiano* aditivo (AWGN).
2. *Modelo del receptor*, que supone que un receptor está compuesto por un filtro pasobanda ideal al que sigue un demodulador ideal apropiado para la aplicación presente; el filtro pasobanda se utiliza para minimizar el efecto del ruido del canal.

Estas suposiciones simplificatorias se establecen para obtener un entendimiento básico de la manera en la que el ruido afecta el desempeño de un receptor. Asimismo, ofrecen un marco de referencia para la comparación de los diferentes esquemas de modulación-demodulación de onda continua.

La figura 2.33 presenta el *modelo de receptor ruidoso* que combina las dos suposiciones anteriores. En esta figura, $s(t)$ denota la señal modulada entrante y $w(t)$ representa el ruido del canal. La *señal recibida* está integrada consecuentemente por la suma de $s(t)$ y $w(t)$; ésta es la señal sobre la cual tiene que trabajar el receptor. El *filtro pasobanda* en el modelo de la figura 2.33 representa la acción combinada de filtrado de los amplificadores sintonizados que se utilizan en el receptor real con el fin de amplificar la señal antes de la demodulación. El ancho de banda del filtro pasobanda es apenas lo suficiente para dejar pasar la señal modulada $s(t)$ sin distorsión. En lo referente al *demodulador* en el modelo de la figura 2.33, sus detalles dependen naturalmente del tipo de modulación que se use.

■ RELACIONES SEÑAL A RUIDO: DEFINICIONES BÁSICAS

Denotemos la densidad espectral de potencia del ruido $w(t)$ por medio de $N_0/2$, que se define tanto para frecuencias positivas como negativas; esto es, N_0 *es la potencia promedio del ruido por ancho de banda unitario medido en el extremo frontal del receptor*. Suponemos también que el filtro pasobanda en el modelo del receptor de la figura 2.33 es ideal y que tiene un ancho de banda igual al ancho de banda de transmisión B_T de la señal modulada $s(t)$ y una frecuencia de banda media igual a la frecuencia f_c de la portadora. La última suposición se justifica por la modulación de doble banda lateral-portadora suprimida (DSB-SC), la modulación de amplitud (AM) completa y la modulación de frecuencia (FM); los casos de modulación de banda lateral única (SSB) y de modulación de banda lateral residual (VSB) requieren consideraciones especiales. Al considerar la frecuencia de banda media del filtro pasobanda igual que la frecuencia f_c de la portadora, es posible modelar la densidad espectral de potencia $S_N(f)$ del ruido $n(t)$, que resulta del paso del ruido blanco $w(t)$ por el filtro, como se indica en la figura 2.34. Por lo común, la

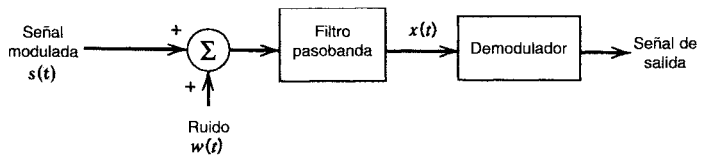


FIGURA 2.33 Modelo de receptor.

frecuencia de portadora f_c es grande en comparación con el ancho de banda de transmisión B_T. Por tanto, podemos tratar el *ruido filtrado* $n(t)$ como un ruido de banda angosta representado en la forma canónica

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \tag{2.79}$$

donde $n_I(t)$ es la *componente de ruido en fase* y $n_Q(t)$ es la *componente de ruido en cuadratura*, medidas ambas con respecto a la onda portadora $A_c\cos(2\pi f_c t)$. La señal filtrada $x(t)$ disponible para la demodulación se define por medio de

$$x(t) = s(t) + n(t) \tag{2.80}$$

Los detalles de $s(t)$ dependen del tipo de modulación que se utiliza. En cualquier evento, la potencia de ruido promedio en la entrada del demodulador es igual al área total bajo la curva de la densidad espectral de potencia $S_N(f)$. De acuerdo con la figura 2.34, vemos fácilmente que esta potencia de ruido promedio es igual a N_0B_T. Dado el formato de $s(t)$, es posible determinar también la potencia de señal promedio en la entrada del demodulador. Con la señal demodulada $s(t)$ y el ruido filtrado $n(t)$ que aparece en forma aditiva en la entrada del demodulador de acuerdo con la ecuación (2.80), podemos continuar y definir una *relación señal a ruido de entrada,* $(\text{SNR})_I$, *como la proporción entre la potencia promedio de la señal modulada* $s(t)$ *y la potencia promedio del ruido filtrado* $n(t)$.

Sin embargo, una medida más útil del desempeño con respecto al ruido es la *relación señal a ruido de salida,* $(\text{SNR})_O$, *definida como la proporción entre la potencia promedio de la señal del mensaje demodulada y la potencia promedio del ruido, ambas medidas en la salida del receptor.* La relación señal de salida a ruido ofrece una medida intuitiva para describir la fidelidad con la que el proceso de demodulación en el receptor recupera la señal del mensaje de la señal modulada en la presencia de ruido aditivo. Para que tal criterio esté bien definido, la señal del mensaje recuperada y la componente de ruido corruptiva deben aparecer *aditivamente* en la salida del demodulador. Esta condición resulta perfectamente válida en el caso de un receptor que utiliza detección coherente. Por otro lado, cuando el receptor emplea la detección de envolvente como en la AM total o la discriminación de frecuencia como en la FM, tenemos que suponer que la potencia promedio del ruido filtrado $n(t)$ es relativamente baja para justificar el uso de una relación señal a ruido de salida como una medida del desempeño del receptor.

La relación de señal a ruido de salida depende, entre otros factores, del tipo de modulación utilizado en el transmisor y del tipo de demodulación que se emplea en el receptor. Por tanto, es informativo comparar las relaciones de señal de salida a ruido de diferentes sistemas de modulación-demodulación. Sin embargo, para que esta comparación sea de utilidad, se debe hacer sobre una base de equidad, como se describe a continuación:

- La señal modulada $s(t)$ transmitida por cada sistema tiene la misma potencia promedio.
- El ruido del canal $w(t)$ tiene la misma potencia promedio media en el ancho de banda W del mensaje.

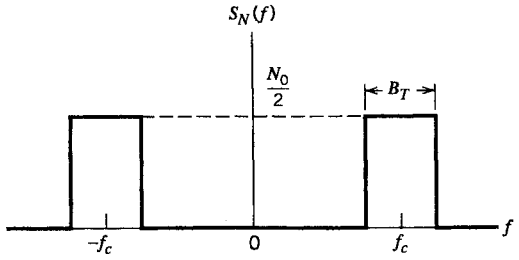


FIGURA 2.34 Característica idealizada del ruido filtrado pasobanda.

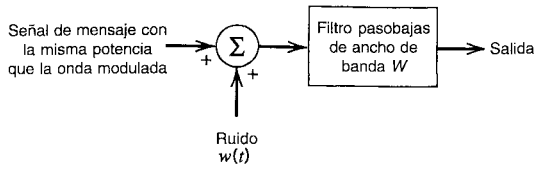


FIGURA 2.35 El modelo de transmisión de banda base, suponiendo una señal de mensaje de banda W, que se utiliza para calcular la relación señal a ruido del canal.

En consecuencia, como un marco de referencia definimos la *relación señal a ruido del canal*, $(\text{SNR})_C$, como *la proporción entre la potencia promedio de la señal modulada y la potencia promedio del ruido del canal en el ancho de banda del mensaje, medidas ambas en la entrada del receptor.* Esta definición se ilustra en la figura 2.35.

Con el fin de comparar los diferentes sistemas de modulación de onda continua, *normalizamos* el desempeño del receptor dividiendo la relación señal a ruido de salida por la relación de señal a ruido del canal. De esa manera, definimos una *figura de mérito* para el receptor de la forma siguiente:

$$\text{Figura de mérito} = \frac{(\text{SNR})_O}{(\text{SNR})_C} \tag{2.81}$$

Naturalmente, cuanto más alto sea el valor de la figura de mérito, tanto mejor será el desempeño del receptor con respecto al ruido. La figura de mérito puede ser igual a uno, menor que uno o mayor que uno, según el tipo de modulación que se utilice, lo cual será manifiesto a partir de la explicación que sigue.

2.11 *Ruido en receptores lineales al utilizar detección coherente*

De las secciones 2.2 y 2.3 recordamos que la demodulación de una onda modulada en amplitud depende de si la portadora está o no suprimida. Cuando se suprime la portadora solemos requerir del uso de detección coherente, en cuyo caso el receptor es *lineal*. De otro modo, cuando la modulación de amplitud incluye la transmisión de la portadora, la demodulación se lleva a cabo utilizando simplemente un detector de envolvente, situación en la que el receptor es *no lineal*. En esta sección estudiamos el efecto del ruido en el desempeño de un receptor lineal. El caso más difícil de un receptor no lineal se deja para la sección 2.12.

Considere el caso de modulación de DSB-SC. La figura 2.36 muestra el modelo de un receptor de DSB-SC que utiliza detección coherente. El empleo de esta última requiere la multiplicación de la señal filtrada $x(t)$ por una onda senoidal $\cos(2\pi f_c t)$ generada localmente y pasar después el producto por un filtro pasobajas. Para simplificar el análisis, suponemos que la amplitud de la onda senoidal que se genera en forma local es unitaria. Sin embargo, para que este esquema de demodulación opere en forma satisfactoria, resulta necesario que el oscilador local se sincronice tanto en fase como en frecuencia con el oscilador que genera la onda portadora en el transmisor. Suponemos que se ha logrado esta sincronización.

La componente de DSB-SC de la señal filtrada $x(t)$ se expresa como

$$s(t) = CA_c \cos(2\pi f_c t) m(t) \tag{2.82}$$

donde $A_c \cos(2\pi f_c t)$ es la onda portadora senoidal y $m(t)$ es la señal del mensaje. En la expresión de $s(t)$ en la ecuación (2.82) hemos incluido un *factor de escalamiento dependiente del sistema* C, cuyo fin es asegurar que la componente de la señal $s(t)$ se mida en las mismas unidades que la componente de ruido aditivo $n(t)$. Suponemos que $m(t)$ es la función de muestreo de un proceso

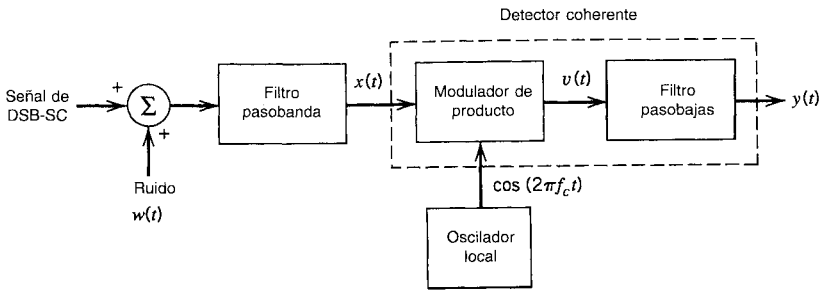


FIGURA 2.36 Modelo de receptor de DSB-SC utilizando detección coherente.

estacionario de media cero, cuya densidad espectral de potencia $S_M(f)$ se limita a una frecuencia máxima W; esto es, W es el *ancho de banda del mensaje*. La potencia promedio P de la señal del mensaje es el área total bajo la curva de la densidad espectral de potencia, como indica

$$P = \int_{-W}^{W} S_M(f)\, df \tag{2.83}$$

La onda portadora es estadísticamente independiente de la señal del mensaje. Para subrayar esta independencia, la portadora debe incluir una fase aleatoria que se distribuye uniformemente sobre 2π radianes. En la ecuación de definición para $s(t)$ este ángulo de fase aleatoria se ha omitido por conveniencia en la presentación. Utilizando el resultado del ejemplo 1.7 del capítulo 1 relativo a procesos aleatorios modulados, es posible expresar la potencia promedio de la componente de la señal modulada de DSB-SC $s(t)$ como $C^2A_c^2P/2$. Con una densidad espectral de ruido de $N_0/2$, la potencia de ruido promedio en el ancho de banda del mensaje W es igual a WN_0. La relación señal a ruido del canal del sistema de modulación de DSB-SC es entonces

$$(\text{SNR})_{C,DSB} = \frac{C^2A_c^2P}{2WN_0} \tag{2.84}$$

donde la constante C^2 en el numerador asegura que este cociente es adimensional.

Después de esto, deseamos determinar la relación señal a ruido de salida del sistema. Utilizando la representación de banda angosta del ruido filtrado $n(t)$, la señal total en la entrada del detector coherente puede expresarse como

$$\begin{aligned} x(t) &= s(t) + n(t) \\ &= CA_c \cos(2\pi f_c t)m(t) + n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \end{aligned} \tag{2.85}$$

donde $n_I(t)$ y $n_Q(t)$ son las componentes en fase y en cuadratura de $n(t)$ con respecto a la portadora. La salida del componente del modulador de producto del detector coherente es consecuentemente

$$\begin{aligned} v(t) &= x(t)\cos(2\pi f_c t) \\ &= \tfrac{1}{2}CA_c m(t) + \tfrac{1}{2}n_I(t) \\ &\quad + \tfrac{1}{2}[CA_c m(t) + n_I(t)]\cos(4\pi f_c t) - \tfrac{1}{2}n_Q(t)\,\text{sen}(4\pi f_c t) \end{aligned}$$

El filtro pasobajas en el detector coherente de la figura 2.36 elimina las componentes de alta frecuencia de $v(t)$, produciéndose la salida del receptor

$$y(t) = \tfrac{1}{2}CA_c m(t) + \tfrac{1}{2}n_I(t) \tag{2.86}$$

La ecuación (2.86) indica lo siguiente:

1. La señal del mensaje $m(t)$ y la componente de ruido en fase $n_I(t)$ del ruido filtrado $n(t)$ aparece aditivamente en la salida del receptor.

2. El detector coherente rechaza por completo la componente en cuadratura $n_Q(t)$ del ruido $n(t)$.

Estos dos resultados son independientes de la relación de señal de entrada a ruido. De tal modo, la detección coherente se distingue por sí sola de otras técnicas de demodulación en una propiedad importante: la componente del mensaje de salida no se mutila y la componente de ruido aparece siempre aditivamente con el mensaje, sin que importe la relación señal a ruido de entrada.

La componente de la señal del mensaje en la salida del receptor es $CA_c m(t)/2$. Por tanto, la potencia promedio de esta componente se expresa como $C^2A_c^2P/4$, donde P es la potencia promedio de la señal del mensaje original $m(t)$ y C es el factor de escalamiento dependiente del sistema al que se hizo referencia antes.

En el caso de la modulación de DSB-SC, el filtro pasobanda en la figura 2.36 tiene un ancho de banda B_T igual a $2W$ para adaptar las bandas laterales superior e inferior de la señal modulada $s(t)$. Por tanto, se concluye que la potencia promedio del ruido filtrado $n(t)$ es igual a $2WN_0$. A partir de la discusión del ruido de banda angosta que se presentó en la sección 1.11, sabemos que la potencia promedio de la componente de ruido en fase (pasobajas) $n_I(t)$ es la misma que la del ruido filtrado (pasobanda) $n(t)$. Puesto que de acuerdo con la ecuación (2.86) la componente de ruido en la salida del receptor es $n_I(t)/2$, se desprende que la potencia promedio del ruido en la salida del receptor es

$$\left(\tfrac{1}{2}\right)^2 2WN_0 = \tfrac{1}{2}WN_0$$

La relación señal a ruido de salida para un receptor de DSB-SC utilizando detección coherente es entonces

$$\begin{aligned}(\text{SNR})_{O,\,DSB\text{-}SC} &= \frac{C^2A_c^2P/4}{WN_0/2} \\ &= \frac{C^2A_c^2P}{2WN_0}\end{aligned} \tag{2.87}$$

Utilizando las ecuaciones (2.84) y (2.87), obtenemos la figura de mérito

$$\left.\frac{(\text{SNR})_O}{(\text{SNR})_C}\right|_{\text{DSB-SC}} = 1 \tag{2.88}$$

Advierta que el factor C^2 es común en ambas relaciones de señal a ruido de la salida y del canal y, por tanto, se cancela al evaluar la figura de mérito.

Al seguir con el análisis de ruido de un detector coherente para SSB, encontramos que, a pesar de las diferencias fundamentales entre él y el detector coherente para modulación de DSB-SC, la figura de mérito es exactamente la misma para ambos; vea el problema 2.49.

Son dos las conclusiones importantes a las que se llega luego de la discusión presentada en esta sección y en el problema 2.49:

1. Para la misma potencia de la señal promedio transmitida o modulada y la misma potencia promedio del ruido en el ancho de banda del mensaje, un receptor de SSB coherente tendrá exactamente la misma relación señal a ruido de salida que un receptor de DSB-SC coherente.
2. En ambos casos, el desempeño del receptor en relación con ruido es exactamente el mismo que el que se obtiene al transmitir simplemente la señal del mensaje en la presencia del mismo ruido de canal. El único efecto del proceso de modulación es trasladar la señal del mensaje a una banda diferente de frecuencia para facilitar su transmisión por un canal pasobanda.

En términos simples, ni la modulación de DSB-SC ni la de SSB ofrecen los medios para un compromiso entre el desempeño ante el ruido mejorado y el ancho de banda del canal aumentado. Esto es un serio problema cuando es requisito indispensable una alta calidad de recepción.

2.12 *Ruido en receptores de AM al utilizar detección de envolvente*

El siguiente análisis de ruido que llevamos a cabo corresponde a un sistema de modulación de amplitud (AM) que usa un detector de envolvente en el receptor, como se muestra en el modelo de la figura 2.37. En una señal de AM completa se transmiten ambas bandas laterales y la onda portadora, de acuerdo a como indica

$$s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t) \tag{2.89}$$

donde $A_c \cos(2\pi f_c t)$ es la onda portadora, $m(t)$ es la señal de mensaje y k_a es una constante que determina el porcentaje de modulación. En la expresión para la componente de la señal modulada en amplitud $s(t)$ dada en la ecuación (2.89), no vemos la necesidad de utilizar un factor de escalamiento, ya que resulta razonable suponer que la amplitud de la portadora A_c tiene las mismas unidades que la componente de ruido aditiva.

La potencia promedio de la componente de la portadora en la señal de AM $s(t)$ es $A_c^2/2$. La potencia promedio de la componente $A_c k_a m(t) \cos(2\pi f_c t)$ que porta la información es $A_c^2 k_a^2 P/2$, donde P es la potencia promedio de la señal del mensaje $m(t)$. La potencia promedio de la señal de AM completa $s(t)$ es, en consecuencia, igual a $A_c^2(1 + k_a^2 P)/2$. Como en el caso de un sistema de DSB-SC, la potencia promedio del ruido en el ancho de banda del mensaje es WN_0. La relación señal a ruido del canal para la AM es entonces

$$(\text{SNR})_{C,AM} = \frac{A_c^2(1 + k_a^2 P)}{2WN_0} \tag{2.90}$$

Para evaluar la relación señal a ruido de salida presentamos primero el ruido filtrado $n(t)$ en términos de sus componentes en fase y en cuadratura. Es posible entonces definir la señal filtrada $x(t)$ que se aplica al detector de envolvente en el modelo del receptor de la figura 2.37 de la manera siguiente:

$$\begin{aligned} x(t) &= s(t) + n(t) \\ &= [A_c + A_c k_a m(t) + n_I(t)]\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \end{aligned} \tag{2.91}$$

Es informativo representar las componentes que integran la señal $x(t)$ por medio de fasores, como en la figura 2.38*a*. De acuerdo con este diagrama fasorial, la salida del receptor se obtiene fácilmente como

$$\begin{aligned} y(t) &= \text{envolvente de } x(t) \\ &= \left\{[A_c + A_c k_a m(t) + n_I(t)]^2 + n_Q^2(t)\right\}^{1/2} \end{aligned} \tag{2.92}$$

La señal $y(t)$ define la salida de un detector de envolvente ideal. La fase de $x(t)$ no nos interesa, pues un detector de envolvente ideal es por completo insensible a las variaciones en la fase de $x(t)$.

La expresión que define a $y(t)$ es un poco compleja y necesita simplificarse de alguna manera para permitir la obtención de resultados significativos. De modo específico, nos gustaría aproximar la salida $y(t)$ como la suma de un término del mensaje más un término debido al ruido. En general, esto es bastante difícil de conseguir. Sin embargo, cuando la potencia prome-

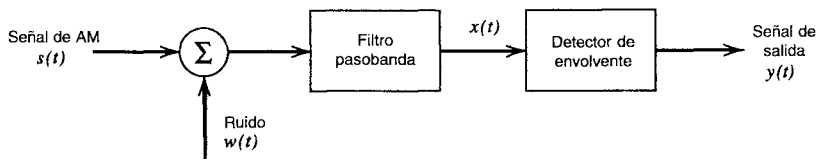


FIGURA 2.37 Modelo de un receptor de AM.

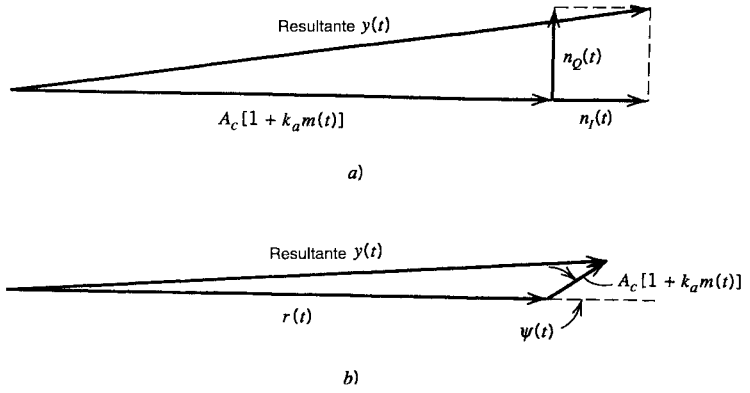


FIGURA 2.38 *a*) Diagrama fasorial de una onda de AM más ruido de banda angosta para el caso de una elevac relación portadora a ruido. *b*) Diagrama fasorial de una onda de AM más ruido de banda angosta para el caso c una baja relación portadora a ruido.

dio de la portadora es grande comparada con la potencia promedio del ruido, de manera que (receptor esté operando en forma satisfactoria, entonces el término $A_c[1 + k_a m(t)]$ de la señal ser grande en comparación con los términos $n_I(t)$ y $n_Q(t)$ del ruido, al menos la mayor parte d(tiempo. En ese caso es posible aproximar la salida $y(t)$ como (vea el problema 2.51):

$$y(t) \simeq A_c + A_c k_a m(t) + n_I(t) \tag{2.93}$$

La presencia del término de cd o constante A_c en la salida del detector de envolvente $y(t$ de la ecuación (2.93) se debe a la demodulación de la onda portadora transmitida. Sin embargo es posible ignorar este término, ya que no guarda relación ninguna con la señal del mensaj $m(t)$. En cualquier caso, puede eliminarse simplemente por medio de un capacitor de bloque(De esa manera, si ignoramos el término A_c de cd en la ecuación (2.93), encontramos que (resto tiene, salvo por factores de escalamiento, una forma similar a la salida de un receptor d DSB-SC utilizando detección coherente. En consecuencia, la relación señal a ruido de salida d un receptor de AM empleando un detector de envolvente es aproximadamente

$$(\text{SNR})_{O,\text{AM}} \simeq \frac{A_c^2 k_a^2 P}{2WN_0} \tag{2.94}$$

Sin embargo, la ecuación (2.94) es válida sólo si se cumplen las dos siguientes condiciones:

1. La potencia promedio del ruido es pequeña en comparación con la potencia promedio de l portadora en la entrada del detector de envolvente.
2. La sensibilidad de la amplitud k_a se ajusta para un porcentaje de modulación menor que igual a 100%.

Utilizando las ecuaciones (2.90) y (2.94), obtenemos la siguiente figura de mérito para la modu lación de amplitud:

$$\left.\frac{(\text{SNR})_O}{(\text{SNR})_C}\right|_{\text{AM}} \simeq \frac{k_a^2 P}{1 + k_a^2 P} \tag{2.95}$$

De tal modo, en tanto que la figura de mérito de un receptor de DSB-SC o la de un receptor d SSB utilizando detección coherente siempre es la unidad, la correspondiente a un receptor de AM utilizando detección de envolvente es siempre menor que la unidad. En otras palabras, *el desem peño ante el ruido de un receptor de AM completo siempre es inferior al de un receptor de DSB-SC*

Esto se debe al desperdicio de la potencia del transmisor, que resulta de la transmisión de la portadora como una componente de la onda de AM.

EJEMPLO 2.4 Modulación de un solo tono

Considere el caso especial de una onda senoidal de frecuencia f_m y amplitud A_m como la onda moduladora indicada por medio de

$$m(t) = A_m \cos(2\pi f_m t)$$

La onda de AM correspondiente es

$$s(t) = A_c[1 + \mu \cos(2\pi f_m t)]\cos(2\pi f_c t)$$

donde $\mu = k_a A_m$ es el factor de modulación. La potencia promedio de la onda moduladora $m(t)$ es (considerando un resistor de carga de 1 ohm)

$$P = \tfrac{1}{2}A_m^2$$

Por tanto, al utilizar la ecuación (2.95), obtenemos

$$\left.\frac{(\text{SNR})_O}{(\text{SNR})_C}\right|_{\text{AM}} = \frac{\frac{1}{2}k_a^2 A_m^2}{1 + \frac{1}{2}k_a^2 A_m^2} = \frac{\mu^2}{2 + \mu^2} \tag{2.96}$$

Cuando $\mu = 1$, lo que corresponde a 100% de modulación, obtenemos una figura de mérito igual a 1/3. Esto significa que, siendo los demás factores iguales, un sistema de AM (utilizando detección de envolvente) debe transmitir tres veces más tanta potencia promedio que un sistema de portadora suprimida (utilizando detección coherente) para alcanzar la misma calidad de desempeño ante el ruido. ◀

EFECTO DE UMBRAL

Cuando la relación portadora a ruido es pequeña comparada con la unidad, el término de ruido domina y el desempeño del detector de envolvente cambia por completo en relación con el que acaba de describirse. En este caso es más conveniente representar el ruido de banda angosta $n(t)$ en términos de su envolvente $r(t)$ y fase $\psi(t)$:

$$n(t) = r(t)\cos[2\pi f_c t + \psi(t)] \tag{2.97}$$

El diagrama fasorial correspondiente para la entrada del detector $x(t) = s(t) + n(t)$ se ilustra en la figura 2.38*b*, donde hemos utilizado la envolvente de ruido como referencia, puesto que ahora es el término dominante. Para el fasor de ruido $r(t)$ se ha agregado un fasor que representa el término $A_c[1 + k_a m(t)]$ de la señal, con el ángulo entre ellos igual a la fase $\psi(t)$ del ruido $n(t)$. En la figura 2.38*b* se ha supuesto que la relación portadora a ruido es tan baja que la amplitud A_c de la portadora es pequeña en comparación con la envolvente del ruido $r(t)$, al menos la mayor parte del tiempo. De tal modo, podemos ignorar la componente en cuadratura de la señal con respecto al ruido y, de esa manera, determinar de acuerdo con la figura 2.38*b* que la salida del detector de envolvente es

$$y(t) \approx r(t) + A_c \cos[\psi(t)] + A_c k_a m(t)\cos[\psi(t)] \tag{2.98}$$

Esta relación revela que cuando la relación portadora a ruido es baja, la salida del detector no tiene componente estrictamente proporcional a la señal del mensaje $m(t)$. El último término de la expresión que define a $y(t)$ contiene la señal del mensaje $m(t)$ multiplicada por el ruido en la forma de $\cos[\psi(t)]$. De la sección 1.11 recordamos que la fase $\psi(t)$ del ruido de banda angosta $n(t)$ se

distribuye de manera uniforme sobre 2π radianes. En consecuencia, se concluye que tenemos una pérdida completa de información, ya que la salida del detector no contiene en absoluto la señal del mensaje $m(t)$. La pérdida de un mensaje en un detector de envolvente que opera a una relación portadora a ruido baja se conoce como el *efecto de umbral*. Por *umbral* queremos decir *un valor de la relación portadora a ruido debajo del cual el desempeño ante el ruido de un detector se deteriora mucho más rápido que proporcionalmente a la relación portadora a ruido*. Es importante reconocer que todo detector no lineal (por ejemplo, el detector de envolvente), exhibe un efecto de umbral. Por otra parte, un efecto de este tipo *no* surge en un detector coherente.

Un análisis matemático riguroso del efecto de umbral para el caso general de una onda de AM está más allá del objetivo de este libro. En la siguiente subsección simplificamos las cosas considerando el caso de una portadora no modulada. A pesar de esta simplificación, aún podemos obtener un gran conocimiento del efecto de umbral que experimenta un detector de envolvente.

*Fórmula general para $(SNR)_0$ en la detección de envolvente**

Considere un detector de envolvente cuya señal de entrada está definida por

$$x(t) = A_c \cos(2\pi f_c t) + n(t) \tag{2.99}$$

donde $A_c \cos(2\pi f_c t)$ es la portadora no modulada y $n(t)$ es la función muestra de ruido blanco gaussiano $N(t)$ de media cero y limitado en banda. La densidad espectral de potencia de $N(t)$ es

$$S_N(t) = \begin{cases} \dfrac{N_0}{2} & \text{para } |f - f_c| \le W \\ 0 & \text{en otro caso} \end{cases} \tag{2.100}$$

Al representar el ruido de banda angosta $n(t)$ en términos de su componente en fase $n_I(t)$ y su componente en cuadratura $n_Q(t)$, es posible expresar la señal de ruido en la entrada del detector como

$$x(t) = (A_c + n_I(t)) \cos(2\pi f_c t) - n_Q(t) \operatorname{sen}(2\pi f_c t) \tag{2.101}$$

Las componentes de ruido $n_I(t)$ y $n_Q(t)$ son procesos aleatorios pasobajas mutuamente independientes, de media cero y gaussianos conjuntamente con densidades espectrales de potencia idénticas [vea la ecuación (1.101)]:

$$S_{N_I}(f) = S_{N_Q}(f) = \begin{cases} S_N(f - f_c) + S_N(f + f_c) & \text{para } |f| \le W \\ 0 & \text{en otro caso} \end{cases} \tag{2.102}$$

Para el problema presente, la señal de entrada consiste en una portadora no modulada con potencia promedio igual a $A_c^2/2$. La potencia promedio del ruido en la entrada del detector es

$$\sigma_N^2 = 2WN_0 \tag{2.103}$$

La *relación portadora a ruido* se define entonces por medio de

$$\begin{aligned} \rho &= \frac{A_c^2/2}{\sigma_N^2} \\ &= \frac{A_c^2}{4WN_0} \end{aligned} \tag{2.104}$$

Es posible considerar a ρ como una relación de señal a ruido de entrada para el problema descrito aquí.

---

*Si al lector no le interesan los detalles matemáticos relativos a cómo afecta el ruido a la detección de envolvente de una señal de AM, puede saltar hasta la ecuación (2.124) y leer los dos casos límite de la fórmula en esa ecuación.

Sin embargo, la determinación de la relación señal a ruido de salida resulta más difícil de llevar a cabo debido a que la salida del detector de envolvente

$$y(t) = \sqrt{(A_c + n_I(t))^2 + n_Q^2(t)} \tag{2.105}$$

es una combinación no lineal de los términos de la señal y el ruido. Sin una separación de corte clara entre la señal y el ruido en la salida del detector $y(t)$, ¿cómo aislamos entonces la contribución de la señal $s(t)$ a $y(t)$ de la contribución debida a ruido $n(t)$? Para resolver este asunto, adoptamos un procedimiento *heurístico* basado en la promediación de la señal. De modo específico, introducimos las siguientes dos definiciones:

1. La *señal de salida media*, s_o, es la diferencia entre la esperanza de $y(t)$ en la presencia combinada de la señal y el ruido, y la esperanza de $y(t)$ en la presencia únicamente del ruido, como se indica por medio de
$$s_o = E[y(t)] - E[y_o(t)] \tag{2.106}$$
donde $y(t)$ se define con base en la ecuación (2.105) y $y_o(t)$, mediante
$$y_o(t) = \sqrt{n_I^2(t) + n_Q^2(t)} \tag{2.107}$$
2. La *potencia media del ruido de salida* es la diferencia entre el valor cuadrático medio de la salida del detector $y(t)$ y el cuadrado del valor medio de $y(t)$, según indica
$$\text{var}[y(t)] = E[y^2(t)] - (E[y(t)])^2 \tag{2.108}$$

Con base en lo anterior, definimos la *relación señal a ruido de salida* como

$$(\text{SNR})_O = \frac{s_o^2}{\text{var}[y(t)]} \tag{2.109}$$

De la sección 1.12 recordamos que la salida del detector de envolvente debida exclusivamente al ruido tiene una distribución de Rayleigh, esto es

$$f_{Y_o}(y) = \begin{cases} \dfrac{y}{\sigma_N^2} \exp\left(-\dfrac{y^2}{2\sigma_N^2}\right), & y \geq 0 \\ 0, & \text{en otro caso} \end{cases} \tag{2.110}$$

La expectación de $y_o(t)$ es entonces

$$\begin{aligned} E[y_o(t)] &= \int_{-\infty}^{\infty} y f_{Y_o}(y)\, dy \\ &= \int_0^{\infty} \frac{y^2}{\sigma_N^2} \exp\left(-\frac{y^2}{2\sigma_N^2}\right) dy \end{aligned} \tag{2.111}$$

A partir de la definición de la *función gamma* para valores reales positivos del argumento x, tenemos

$$\Gamma(x) = \int_0^{\infty} z^{x-1} \exp(-z)\, dz \tag{2.112}$$

Por tanto, es posible reescribir la ecuación (2.111) como

$$\begin{aligned} E[y_o(t)] &= \sqrt{2}\sigma_N \Gamma\left(\frac{3}{2}\right) \\ &= \sqrt{\frac{\pi}{2}}\sigma_N \end{aligned} \tag{2.113}$$

donde hemos utilizado el valor $\Gamma(3/2) = \sqrt{\pi/2}$. Para calcular la señal media s_o en la salida del detector, también necesitamos la esperanza de $y(t)$. En virtud de la presencia combinada de la señal y el ruido, recordamos de la sección 1.13 que $y(t)$ presenta una distribución tipo Rician, como indica

$$f_Y(y) = \begin{cases} \dfrac{y}{\sigma_N^2}\exp\left(-\dfrac{y^2 + A_c^2}{2\sigma_N^2}\right) I_0\left(\dfrac{A_c y}{\sigma_N^2}\right) & \text{para } y \geq 0 \\ 0 & \text{en otro caso} \end{cases} \tag{2.114}$$

donde $I_0(\cdot)$ es la función modificada de Bessel de primer tipo de orden cero (vea el apéndice 3). Por tanto,

$$E[y(t)] = \int_0^\infty \frac{y^2}{\sigma_N^2}\exp\left(-\frac{y^2 + A_c^2}{2\sigma_N^2}\right) I_0\left(\frac{A_c y}{\sigma_N^2}\right) dy \tag{2.115}$$

Poniendo $A_c y/\sigma_N^2 = u$ y reconociendo que $\rho = A_c^2/2\sigma_N^2$, podemos arreglar esta esperanza en la forma

$$E[y(t)] = \frac{\sigma_N}{(2\rho)^{3/2}}\exp(-\rho)\int_0^\infty u^2 \exp\left(-\frac{u^2}{4\rho}\right) I_0(u)\, du \tag{2.116}$$

La integral en la ecuación (2.16) puede escribirse en una forma concisa utilizando *funciones hipergeométricas confluentes*; vea el apéndice 4. En particular, al utilizar la representación integral

$$\int_0^\infty u^{m-1}\exp(-b^2u^2) I_0(u)\, du = \frac{\Gamma(m/2)}{2b^m}\left({}_1F_1\left(\frac{m}{2};1;\frac{1}{4b^2}\right)\right) \tag{2.117}$$

con $m = 3$, $\Gamma(m/2) = \sqrt{\pi/2}$ y $b^2 = 1/4\rho$, podemos expresar la esperanza de $y(t)$ en términos de la función hipergeométrica confluente ${}_1F_1(3/2;1;\rho)$ como

$$E[y(t)] = \sqrt{\frac{\pi}{2}}\sigma_N \exp(-\rho)\left({}_1F_1\left(\frac{3}{2};1;\rho\right)\right) \tag{2.118}$$

Es factible simplificar aún más las cosas utilizando la siguiente identidad:

$$\exp(-u)({}_1F_1(\alpha;\beta;u)) = {}_1F_1(\beta - \alpha;\beta;-u) \tag{2.119}$$

y de ese modo expresar finalmente la expresión $y(t)$ en la forma concisa

$$E[y(t)] = \sqrt{\frac{\pi}{2}}\sigma_N\left({}_1F_1\left(-\frac{1}{2};1;-\rho\right)\right) \tag{2.120}$$

Por tanto, el uso de las ecuaciones (2.113) y (2.120) en la ecuación (2.106) produce la señal de la salida media como

$$s_o = \sqrt{\frac{\pi}{2}}\sigma_N\left({}_1F_1\left(-\frac{1}{2};1;-\rho\right) - 1\right) \tag{2.121}$$

cuya dependencia en la desviación estándar σ_N del ruido $n(t)$ es testimonio de la entremezcla de la señal y el ruido a la salida del detector.

Siguiendo un procedimiento similar, podemos expresar el valor cuadrático medio de la salida del detector $y(t)$ como

$$\begin{aligned} E[y^2(t)] &= \int_0^\infty \frac{y^3}{\sigma_N^2}\exp\left(-\frac{y^2 + A_c^2}{2\sigma_N^2}\right) I_0\left(\frac{A_c y}{\sigma_N^2}\right) dy \\ &= 2\sigma_N^2({}_1F_1(-1;1;-\rho)) \end{aligned} \tag{2.122}$$

Por tanto, al usar las ecuaciones (2.120) y (2.122) en la (2.108), se produce la potencia media del ruido de salida como

$$\text{var}[y(t)] = 2\sigma_N^2\left({}_1F_1(-1;1;-\rho) - \frac{\pi}{4}\left({}_1F_1\left(-\frac{1}{2};1;-\rho\right)\right)^2\right) \tag{2.123}$$

Por último, el uso de las ecuaciones (2.121) y (2.123) en la (2.109) origina la relación señal a ruido de salida para el problema de la detección de envolvente que estamos considerando como

$$(\text{SNR})_O = \frac{\left({}_1F_1\left(-\frac{1}{2};1;-\rho\right)-1\right)^2}{(4/\pi)({}_1F_1(-1;1;-\rho)) - \left({}_1F_1\left(-\frac{1}{2};1;-\rho\right)\right)^2} \tag{2.124}$$

La ecuación (2.124) es la fórmula general para la señal a ruido de salida de un detector de envolvente cuya entrada está compuesta por una portadora no modulada y ruido blanco gaussiano limitado en banda. Dos casos límite de esta fórmula general resultan de particular interés:

1. *Relación portadora a ruido elevada.* Para un valor considerable de ρ, es posible utilizar la siguiente fórmula asintótica (vea el apéndice 4)

$$ {}_1F_1\left(-\frac{1}{2};1;-\rho\right) \simeq 2\sqrt{\frac{\rho}{\pi}} \quad \text{para } \rho \to \infty \tag{2.125}$$

Además, la siguiente identidad

$$ {}_1F_1(-1;1;-\rho) = 1+\rho \tag{2.126}$$

se cumple exactamente para toda ρ. En consecuencia, el uso de las ecuaciones (2.125) y (2.126) en la (2.124) produce la siguiente fórmula aproximada para la relación señal a ruido de salida:

$$(\text{SNR})_O \simeq \rho \quad \text{para } \rho \to \infty \tag{2.127}$$

donde hemos ignorado las contribuciones de $\rho^{1/2}$ y de ρ^0 en el numerador de la ecuación (2.124) como las subdominantes en comparación con ρ para un valor grande de esta misma. La ecuación (2.127) muestra que para una ρ de portadora a ruido elevada, el detector de envolvente se comporta como un detector coherente, ya que la relación señal a ruido de salida es proporcional a la relación señal a ruido de entrada.

2. *Relación portadora a ruido pequeña.* Para ρ pequeña, tenemos (vea el apéndice 4)

$$ {}_1F_1(a;c;-\rho) \simeq 1 - \frac{a}{c}\rho \quad \text{para } \rho \to 0 \tag{2.128}$$

Por tanto, utilizando esta fórmula asintótica, es posible aproximar la relación señal a ruido de salida para ρ pequeña como

$$\begin{aligned}(\text{SNR})_O &\simeq \frac{\pi\rho^2}{16-4\pi} \\ &\simeq 0.91\rho^2 \qquad \text{para } \rho \to 0\end{aligned} \tag{2.129}$$

donde, en el denominador, hemos ignorado las contribuciones debidas a ρ y ρ^2 como las subdominantes en comparación con ρ^0 para el caso de ρ pequeña. La ecuación (2.129) muestra que para una relación portadora a ruido pequeña, la relación señal a ruido de salida del detector de envolvente es proporcional a la relación al cuadrado de señal a ruido de salida.

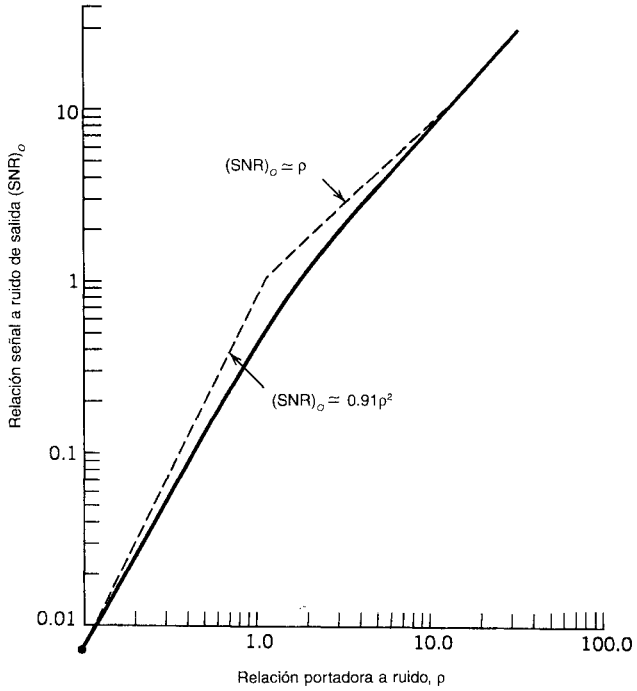


FIGURA 2.39 Relación señal a ruido de salida de un detector de envolvente para una relación variable de portadora a ruido.

Las conclusiones que se extraen de los dos casos límite considerados aquí son que un detector de envolvente favorece las señales fuertes y penaliza las débiles. El fenómeno de las señales débiles que son penalizadas por el detector se conoce como *supresión de señal débil*, el cual es una manifestación del efecto de umbral.

Al utilizar la fórmula de la ecuación (2.124), hemos graficado en la figura 2.39 la relación señal a ruido de salida $(SNR)_O$ del detector de envolvente en función de la relación de portadora a ruido ρ utilizando valores tabulados de las funciones hipergeométricas confluentes. Esta figura incluye también las dos asíntotas para ρ grande y ρ pequeña. A partir de esta figura observamos que la relación señal a ruido de salida se aparta del comportamiento lineal alrededor de una relación de portadora a ruido de 10 dB (esto es $\rho \simeq 10$).

2.13 *Ruido en receptores de FM*

Por último, dirigimos nuestra atención al análisis de ruido de un sistema de modulación de frecuencia (FM), para el cual recurrimos al modelo del receptor que se ilustra en la figura 2.40. Como antes, el ruido $w(t)$ se modela como ruido blanco gaussiano de media cero y densidad espectral de potencia

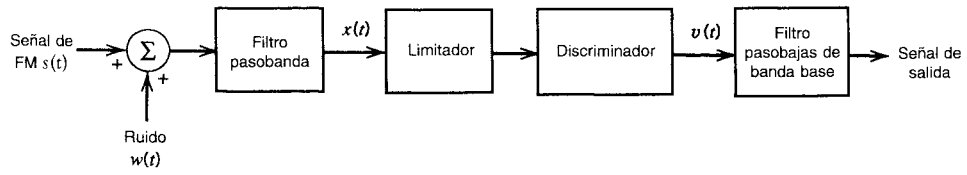


FIGURA 2.40 Modelo de un receptor de FM.

$N_0/2$. La señal de FM recibida $s(t)$ tiene una frecuencia de portadora f_c y un ancho de banda de transmisión B_T, tal que sólo una cantidad ignorable de potencia se encuentra fuera de la banda de frecuencia $f_c \pm B_T/2$ para frecuencias positivas, y similarmente para frecuencias negativas.

Como en el caso de AM, el filtro pasobanda tiene una frecuencia de banda media f_c y un ancho de banda B_T y, en consecuencia, deja pasar la señal de FM casi sin distorsión. En la mayoría de los casos, B_T es pequeña comparada con la frecuencia de banda media f_c, por lo que es posible utilizar la representación de banda angosta para $n(t)$, la versión filtrada del ruido del canal $w(t)$, en términos de sus componentes en fase y en cuadratura.

En un sistema de FM, la señal del mensaje se transmite mediante variaciones de la frecuencia instantánea de una onda portadora senoidal, y su amplitud se mantiene constante. Por tanto, cualesquiera variaciones de la amplitud de la portadora en la entrada del receptor deben resultar del ruido o de la interferencia. El *limitador* de amplitud, que sigue al filtro pasobanda en el modelo del receptor de la figura 2.40, se emplea para eliminar las variaciones de amplitud sujetando la onda modulada a la salida del filtro casi en el eje cero. La onda rectangular que se produce se redondea mediante otro filtro pasobanda que es parte integral del limitador, suprimiéndose de esa manera los armónicos de la frecuencia de la portadora. Por tanto, la salida del filtro es otra vez senoidal, con una amplitud prácticamente independiente de la amplitud de la portadora en la entrada del receptor.

El discriminador en el modelo de la figura 2.40 consta de dos componentes:

1. Una *red de pendiente o diferenciador* como una respuesta en frecuencia imaginaria que varía linealmente con la frecuencia. Ésta produce una onda modulada híbrida en la que varían tanto la amplitud como la frecuencia de acuerdo con la señal del mensaje.
2. Un detector de envolvente que recupera la variación de amplitud y de ese modo reproduce la señal del mensaje.

La red de pendiente y el detector de envolvente suelen ponerse en práctica como partes integrales de una sola unidad física.

El *filtro posdetección*, denominado "filtro pasobanda de banda base" en la figura 2.40, tiene un ancho de banda que es apenas lo bastante grande para ajustar la componente de frecuencia más alta de la señal del mensaje. Este filtro elimina las componentes fuera de banda del ruido en la salida del discriminador y mantiene de ese modo en un mínimo el efecto del ruido de salida.

El ruido filtrado $n(t)$ a la salida del filtro pasobanda en la figura 2.40 se define en términos de sus componentes en fase y en cuadratura por medio de

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\mathrm{sen}(2\pi f_c t)$$

De modo equivalente, podemos expresar $n(t)$ en términos de su envolvente y fase como

$$n(t) = r(t)\cos[(2\pi f_c t) + \psi(t)] \tag{2.130}$$

donde la envolvente es

$$r(t) = \left[n_I^2(t) + n_Q^2(t)\right]^{1/2} \tag{2.131}$$

y la fase corresponde a

$$\psi(t) = \tan^{-1}\left[\frac{n_Q(t)}{n_I(t)}\right] \tag{2.132}$$

La envolvente $r(t)$ tiene una distribución de Rayleigh, y la fase $\psi(t)$ se distribuye de manera uniforme sobre 2π radianes (vea la sección 1.12).

La señal de FM entrante $s(t)$ está definida por

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\,d\tau\right] \tag{2.133}$$

donde A_c es la amplitud de la portadora, f_c es la frecuencia de la portadora, k_f corresponde a la sensibilidad de la frecuencia y $m(t)$ es la señal del mensaje. Advierta que, como en el caso de la AM estándar, en la FM no es necesario introducir un factor de escalamiento en la definición de la señal modulada $s(t)$, puesto que resulta razonable suponer que su amplitud A_c tiene las mismas unidades que la componente de ruido aditiva $n(t)$. Para proceder, definimos

$$\phi(t) = 2\pi k_f \int_0^t m(\tau)\, d\tau \tag{2.134}$$

Por tanto, podemos expresar $s(t)$ en la forma simple

$$s(t) = A_c \cos[2\pi f_c t + \phi(t)] \tag{2.135}$$

La señal ruidosa en la salida del filtro pasobanda es entonces

$$\begin{aligned} x(t) &= s(t) + n(t) \\ &= A_c \cos[2\pi f_c t + \phi(t)] + r(t) \cos[2\pi f_c t + \psi(t)] \end{aligned} \tag{2.136}$$

Es informativo representar $x(t)$ mediante un diagrama fasorial, como en la figura 2.41. En este diagrama hemos utilizado el término de la señal como referencia. La fase $\theta(t)$ del fasor resultante que representa a $x(t)$ se obtiene directamente de la figura 2.41 como

$$\theta(t) = \phi(t) + \tan^{-1}\left\{\frac{r(t)\,\text{sen}[\psi(t) - \phi(t)]}{A_c + r(t)\cos[\psi(t) - \phi(t)]}\right\} \tag{2.137}$$

La envolvente de $x(t)$ no nos interesa, debido a que cualesquiera variaciones de la envolvente en la salida del filtro pasobanda las elimina el limitador.

Nuestra motivación radica en determinar el error en la frecuencia instantánea de la onda portadora provocado por la presencia del ruido filtrado $n(t)$. Con la suposición de que el discriminador es ideal, su salida es proporcional a $\theta'(t)/2\pi$, donde $\theta'(t)$ es la derivada de $\theta(t)$ con respecto al tiempo. Sin embargo, en vista de la complejidad de la expresión que define a $\theta(t)$, necesitamos hacer ciertas aproximaciones para simplificar, de manera que nuestro análisis pueda producir resultados útiles.

Suponemos que la relación de portadora a ruido medida a la entrada del discriminador es grande en comparación con la unidad. Dejamos que R denote la variable aleatoria que se obtiene al observar (en algún tiempo fijo) el proceso de la envolvente con la función de muestreo $r(t)$ [resultado del ruido $n(t)$] . Entonces, al menos la mayor parte del tiempo, la variable aleatoria R es pequeña comparada con la amplitud A_c de la portadora, y por ello la expresión de la fase $\theta(t)$ se simplifica considerablemente:

$$\theta(t) \simeq \phi(t) + \frac{r(t)}{A_c}\,\text{sen}[\psi(t) - \phi(t)] \tag{2.138}$$

o, recurriendo a la expresión para $\phi(t)$ dada en la ecuación (2.134),

$$\theta(t) \simeq 2\pi k_f \int_0^t m(\tau)\, d\tau + \frac{r(t)}{A_c}\,\text{sen}[\psi(t) - \phi(t)] \tag{2.139}$$

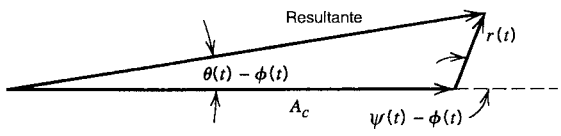


FIGURA 2.41 Diagrama fasorial para una onda FM más ruido de banda angosta para el caso de una elevada relación portadora a ruido.

La salida del discriminador es entonces

$$\begin{aligned} v(t) &= \frac{1}{2\pi}\frac{d\theta(t)}{dt} \\ &= k_f m(t) + n_d(t) \end{aligned} \tag{2.140}$$

donde el término del ruido $n_d(t)$ está definido por

$$n_d(t) = \frac{1}{2\pi A_c}\frac{d}{dt}\{r(t)\,\text{sen}[\psi(t) - \phi(t)]\} \tag{2.141}$$

De este modo, vemos que siempre que es alta la relación de portadora a ruido, la salida del discriminador $v(t)$ está compuesta por la señal del mensaje original $m(t)$ multiplicada por el factor constante k_f, más una componente de ruido aditiva $n_d(t)$. En consecuencia, podemos utilizar la relación señal a ruido de salida que se definió antes para evaluar la calidad del desempeño del receptor de FM. Sin embargo, antes de hacerlo es instructivo ver si es posible simplificar la expresión que define al ruido $n_d(t)$.

De acuerdo con el diagrama fasorial de la figura 2.41 observamos que el efecto de las variaciones en la fase $\psi(t)$ del ruido de banda angosta aparece referido al término de la señal $\phi(t)$. Sabemos que la fase $\psi(t)$ se distribuye uniformemente sobre 2π radianes. Por tanto, sería tentador suponer que la diferencia de fase $\psi(t) - \phi(t)$ también se distribuye de manera uniforme sobre 2π radianes. Si tal suposición fuera cierta, entonces el ruido $n_d(t)$ en la salida del discriminador sería independiente de la señal moduladora y dependería sólo de las características de la portadora y del ruido de banda angosta. Las consideraciones teóricas demuestran que esta suposición se justifica siempre y cuando sea alta la relación portadora a ruido.[9] En ese caso podemos simplificar la ecuación (2.141) como:

$$n_d(t) \simeq \frac{1}{2\pi A_c}\frac{d}{dt}\{r(t)\,\text{sen}[\psi(t)]\} \tag{2.142}$$

Sin embargo, de las ecuaciones de definición para $r(t)$ y $\psi(t)$, notamos que la componente en cuadratura $n_Q(t)$ del ruido filtrado $n(t)$ es

$$n_Q(t) = r(t)\,\text{sen}[\psi(t)] \tag{2.143}$$

Por tanto, la ecuación (2.142) puede escribirse como

$$n_d(t) \simeq \frac{1}{2\pi A_c}\frac{dn_Q(t)}{dt} \tag{2.144}$$

Esto significa que *el ruido aditivo* $n_d(t)$ *que aparece a la salida del discriminador se determina de manera efectiva por medio de la amplitud* A_c *de la portadora y por la componente en cuadratura* $n_Q(t)$ *del ruido de banda angosta* $n(t)$.

La relación señal a ruido de salida se define como la proporción entre la potencia promedio de la señal de salida y la potencia promedio del ruido de salida. De acuerdo con la ecuación (2.140), vemos que la componente del mensaje en la salida del discriminador y, consecuentemente la salida del filtro pasobajas, es $k_f m(t)$. Por tanto, la potencia promedio de la señal de salida es igual a $k_f^2 P$, donde P es la potencia promedio de la señal del mensaje $m(t)$.

Para determinar la potencia promedio del ruido de salida, advertimos que el ruido $n_d(t)$ en la salida del discriminador es proporcional a la derivada en el tiempo de la componente del ruido en cuadratura $n_Q(t)$. Puesto que la diferenciación de una función con respecto al tiempo corresponde a la multiplicación de su transformada de Fourier por $j2\pi f$, se desprende que es posible obtener el procesamiento del ruido $n_d(t)$ al hacer pasar $n_Q(t)$ por un filtro lineal con respuesta en frecuencia igual a

$$\frac{j2\pi f}{2\pi A_c} = \frac{jf}{A_c}$$

Esto quiere decir que la densidad espectral de potencia $S_{N_d}(f)$ del ruido $n_d(t)$ se relaciona con la densidad espectral de potencia $S_{N_Q}(f)$ de la componente de ruido en cuadratura $n_Q(t)$ de la manera siguiente:

$$S_{N_d}(f) = \frac{f^2}{A_c^2} S_{N_Q}(f) \tag{2.145}$$

Con el filtro pasobanda en el modelo de receptor de la figura 2.40 cuya respuesta en frecuencia ideal está caracterizada por el ancho de banda B_T y la frecuencia de banda media f_c, se concluye que el ruido de banda angosta $n(t)$ tendrá una característica de densidad espectral de potencia que es de forma similar. Esto implica que la componente en cuadratura $n_Q(t)$ del ruido de banda angosta $n(t)$ tendrá la característica pasobajas ideal que se muestra en la figura 2.42*a*. La correspondiente densidad espectral de potencia del ruido $n_d(t)$ se presenta en la figura 2.42*b*; esto es,

$$S_{N_d}(f) = \begin{cases} \dfrac{N_0 f^2}{A_c^2}, & |f| \le \dfrac{B_T}{2} \\ 0, & \text{en otro caso} \end{cases} \tag{2.146}$$

En el modelo de receptor de la figura 2.40, a la salida del discriminador sigue un filtro pasobajas con un ancho de banda igual al ancho de banda del mensaje W. Para la FM de banda ancha, encontramos casi siempre que W es más pequeña que $B_T/2$, donde B_T es el ancho de banda de transmisión de la señal de FM. Esto quiere decir que las componentes de ruido fuera de banda $n_d(t)$ serán rechazadas. Por tanto, la densidad espectral de potencia $S_{N_o}(f)$ del ruido $n_o(t)$ que aparece a la salida del receptor está definida por

$$S_{N_o}(f) = \begin{cases} \dfrac{N_0 f^2}{A_c^2}, & |f| \le W \\ 0, & \text{en otro caso} \end{cases} \tag{2.147}$$

como se muestra en la figura 2.42*c*. La potencia promedio del ruido de salida se determina integrando la densidad espectral de potencia $S_{N_o}(f)$ de $-W$ a W. De tal modo obtenemos el siguiente resultado:

$$\begin{aligned} \text{Potencia promedio del ruido de salida} &= \frac{N_0}{A_c^2} \int_{-W}^{W} f^2 \, df \\ &= \frac{2N_0 W^3}{3A_c^2} \end{aligned} \tag{2.148}$$

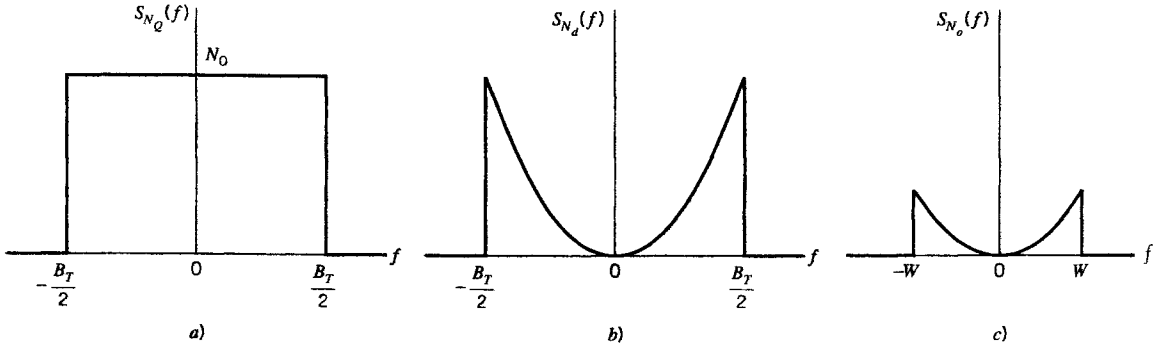


FIGURA 2.42 Análisis de ruido del receptor de FM. *a*) Densidad espectral de potencia de la componente en cuadratura $n_Q(t)$ del ruido de banda angosta $n(t)$. *b*) Densidad espectral de potencia $n_d(t)$ a la salida del discriminador. *c*) Densidad espectral de potencia $n_o(t)$ a la salida del receptor.

Notamos que la potencia promedio del ruido de salida es inversamente proporcional a la potencia promedio $A_c^2/2$ de la portadora. En consecuencia, en un sistema de FM, el incremento de la potencia de la portadora tiene un *efecto de silenciamiento de ruido*.

Determinamos antes la potencia promedio de la señal de salida como $k_f^2 P$. De ese modo, siempre que sea alta la relación de portadora a ruido, podemos dividir esta potencia promedio de la señal de salida por la potencia promedio del ruido de salida de la ecuación (2.148) para obtener la relación señal a ruido de salida

$$(\text{SNR})_{\text{O,FM}} = \frac{3A_c^2 k_f^2 P}{2N_0 W^3} \tag{2.149}$$

La potencia promedio en la señal modulada $s(t)$ es $A_c^2/2$, y la potencia de ruido promedio en el ancho de banda del mensaje es WN_0. Por consiguiente, la relación de señal a ruido del canal es

$$(\text{SNR})_{\text{C,FM}} = \frac{A_c^2}{2WN_0} \tag{2.150}$$

Al dividir la relación señal a ruido de salida entre la relación de señal a ruido del canal, obtenemos la siguiente figura de mérito para la modulación en frecuencia:

$$\left.\frac{(\text{SNR})_{\text{O}}}{(\text{SNR})_{\text{C}}}\right|_{\text{FM}} = \frac{3k_f^2 P}{W^2} \tag{2.151}$$

De la sección 2.7 recordamos que la desviación de frecuencia Δf es proporcional a la sensibilidad de frecuencia k_f del modulador. Asimismo, por definición, la razón de desviación D es igual a la desviación de frecuencia Δf dividida entre el ancho de banda del mensaje W. En otras palabras, la razón de desviación D es proporcional a la razón $k_f P^{1/2}/W$. Se concluye entonces de acuerdo con la ecuación (2.151) que la figura de mérito de un sistema de FM de banda ancha es una función cuadrática de la razón de desviación. Ahora bien, en la FM de banda ancha, el ancho de banda de transmisión B_T es más o menos proporcional a la razón de desviación D. En consecuencia, es posible establecer que *cuando es alta la relación portadora a ruido, un aumento en el ancho de banda de transmisión B_T proporciona un incremento cuadrático correspondiente en la relación señal a ruido de salida o figura de mérito del sistema de FM*. El punto importante que hay que observar en este enunciado es que, a diferencia de la modulación de amplitud, el uso de la modulación de frecuencia ofrece un mecanismo práctico para intercambiar el ancho de transmisión incrementado por un mejor desempeño ante el ruido.

Ejemplo 2.5 Modulación de un solo tono

Considere el caso de una onda senoidal de frecuencia f_m como la señal moduladora y suponga una desviación de frecuencia pico Δf. La señal de FM modulada está definida entonces por

$$s(t) = A_c \cos\left[2\pi f_c t + \frac{\Delta f}{f_m}\operatorname{sen}(2\pi f_m t)\right]$$

Por tanto, podemos escribir

$$2\pi k_f \int_0^t m(\tau)\, d\tau = \frac{\Delta f}{f_m}\operatorname{sen}(2\pi f_m t)$$

Diferenciando ambos miembros con respecto al tiempo y resolviendo para $m(t)$, obtenemos

$$m(t) = \frac{\Delta f}{k_f}\cos(2\pi f_m t)$$

Por tanto, la potencia promedio de la señal del mensaje $m(t)$, que se genera a través de una carga de un 1 ohm, es

$$P = \frac{(\Delta f)^2}{2k_f^2}$$

Sustituyendo este resultado en la fórmula para la relación señal a ruido de salida dada en la ecuación (2.149), obtenemos

$$(\text{SNR})_{O,\text{FM}} = \frac{3A_c^2(\Delta f)^2}{4N_0W^3} = \frac{3A_c^2\beta^2}{4N_0W}$$

donde $\beta = \Delta f/W$ es el índice de modulación. Utilizando la ecuación (2.151) para evaluar la correspondiente figura de mérito, obtenemos

$$\left.\frac{(\text{SNR})_O}{(\text{SNR})_C}\right|_{\text{FM}} = \frac{3}{2}\left(\frac{\Delta f}{W}\right)^2 = \frac{3}{2}\beta^2 \tag{2.152}$$

Es importante advertir que el índice de modulación $\beta = \Delta f/W$ está determinado por el ancho de banda W del filtro pasobajas de posdetección y no con la frecuencia f_m del mensaje senoidal, salvo cuando el filtro suela elegirse de manera que deje pasar el espectro del mensaje deseado; lo anterior es sólo un asunto de diseño consistente. Para un ancho de banda W especificado del sistema, la frecuencia f_m del mensaje senoidal puede encontrarse en cualquier parte entre 0 y W y produciría la misma relación señal a ruido de salida.

Resulta de interés particular comparar el desempeño ante el ruido de los sistemas de AM y FM. Una forma aguda de llevar a cabo esta comparación es considerar las figuras de mérito de los dos sistemas basados en una señal moduladora senoidal. Para un sistema de AM que opere con una señal moduladora senoidal y modulación de 100%, tenemos (del ejemplo 2.4):

$$\left.\frac{(\text{SNR})_O}{(\text{SNR})_C}\right|_{\text{AM}} = \frac{1}{3}$$

Comparando esta figura de mérito con el resultado correspondiente que se describe en la ecuación (2.152) para un sistema de FM, vemos que el uso de la modulación de frecuencia ofrece la posibilidad de mejorar el desempeño ante el ruido sobre la modulación de amplitud cuando

$$\tfrac{3}{2}\beta^2 > \tfrac{1}{3}$$

esto es,

$$\beta > \frac{\sqrt{2}}{3} = 0.471$$

En consecuencia, podemos considerar $\beta = 0.5$ como la definición aproximada de la *transición entre la FM de banda angosta y la FM de banda ancha*. Este enunciado, basado en consideraciones de ruido, confirma además una observación similar que se hizo en la sección 2.7 cuando se consideró el ancho de banda de ondas de FM. ◀

■ EFECTO DE CAPTURA

La capacidad inherente de un sistema de FM para minimizar los efectos de señales indeseables (por ejemplo, ruido, como acaba de explicarse) se aplica también a la *interferencia* producida por otra señal modulada en frecuencia cuyo contenido de frecuencia se acerca a la frecuencia de la portadora de la onda de FM deseada. Sin embargo, la supresión de interferencia en un receptor de FM funciona bien sólo cuando la interferencia es más débil que la entrada de FM que se desea. Cuando la interferencia es la más intensa de una de las dos, el receptor se cierra en la señal más fuerte y por ello suprime la entrada de FM de interés. Cuando son casi de la misma

intensidad, el receptor fluctúa de un lado a otro entre ellos. Este fenómeno se conoce como *efecto de captura*, el cual describe otra característica distintiva de la modulación de frecuencia.

Efecto de umbral de FM

La fórmula de la ecuación (2.149), que define la relación señal a ruido de salida de un receptor de FM, sólo es válida si la relación portadora a ruido, medida a la entrada del discriminador, es elevada en comparación con la unidad. Se encuentra experimentalmente que cuando la potencia del ruido de entrada aumenta de manera que disminuye la relación portadora a ruido, *falla* el receptor de FM. Al principio se escuchan clics individuales en la salida del receptor, y cuando la relación de la portadora al ruido disminuye aún más, los clics surgen rápidamente como parte de un *sonido de crepitación* o de *chisporroteo*. Cerca del punto de falla, la ecuación (2.149) empieza a predecir de manera incorrecta valores de la relación de la señal de salida a ruido mayores que los reales. Este fenómeno se conoce como el *efecto de umbral*.[10] El umbral se define como la relación mínima portadora a ruido que produce una mejora de FM que no es deteriorada de manera importante a partir del valor predicho por la fórmula usual de señal a ruido, suponiendo una potencia de ruido pequeña.

Para una explicación cualitativa del efecto de umbral de FM, considere primero el caso en el que no hay señal presente, de manera que la portadora no está modulada. En ese caso la señal compuesta a la entrada del discriminador de frecuencia es

$$x(t) = [A_c + n_I(t)]\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t) \tag{2.153}$$

donde $n_I(t)$ y $n_Q(t)$ son las componentes en fase y en cuadratura del ruido de banda angosta $n(t)$ con respecto a la onda portadora. El diagrama fasorial de la figura 2.43 presenta la relación de fase entre las diversas componentes de $x(t)$ en la ecuación (2.153). Cuando las amplitudes y fases de $n_I(t)$ y $n_Q(t)$ cambian con el tiempo de una manera aleatoria, el punto P_1 [la punta del fasor que representa a $x(t)$] vaga alrededor del punto P_2 (la punta del fasor que representa a la portadora). Cuando la relación de portadora a ruido es grande, $n_I(t)$ y $n_Q(t)$ suelen ser mucho más pequeñas que la amplitud A_c de la portadora, y de ese modo el punto errático P_1 en la figura 2.43 consume la mayor parte de su tiempo cerca del punto P_2. De ese modo, el ángulo $\theta(t)$ es aproximadamente $n_Q(t)/A_c$ hasta dentro de un múltiplo de 2π. En cambio, cuando la relación portadora a ruido es baja, el punto errático P_1 oscila ocasionalmente por el origen y $\theta(t)$ aumenta o disminuye en 2π radianes. La figura 2.44*a* ilustra cómo de una manera aproximada las excursiones en $\theta(t)$, descritas en la figura 2.44, producen componentes similares a impulsos en $\theta'(t) = d\theta/dt$. La salida del discriminador $v(t)$ es igual a $\theta'(t)/2\pi$. Estos componentes semejantes a impulsos tienen diferentes alturas dependiendo de qué tan cerca el punto errático P_1 se aproxime al origen O, aunque todos tienen áreas casi iguales a $\pm 2\pi$ radianes, como se ilustra en la figura 2.44*b*. Cuando la señal que se muestra en la figura 2.44*b* pasa por el filtro pasobajas de posdetección, los componentes similares a impulsos correspondientes pero más anchos se excitan en la salida del receptor y se escuchan como clics. Estos últimos se producen sólo cuando $\theta(t)$ cambia en $\pm 2\pi$ radianes.

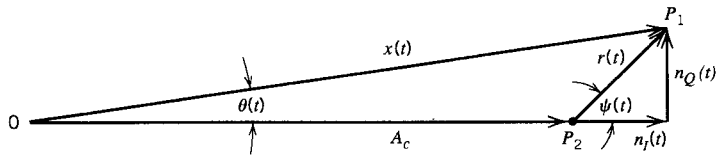


FIGURA 2.43 Interpretación del diagrama fasorial de la ecuación (2.153).

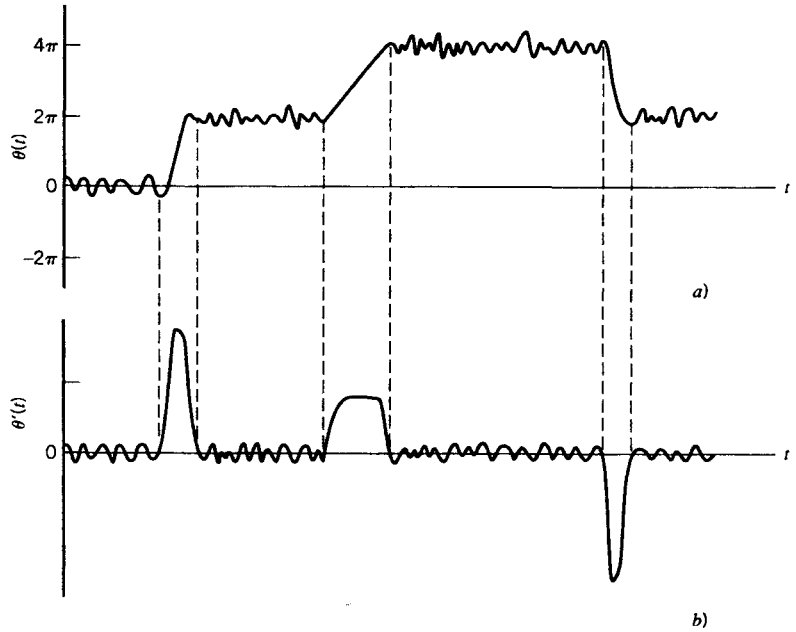


FIGURA 2.44 Ilustración de componentes similares a impulsos en $\theta'(t) = d\theta(t)/dt$ producidos por cambios de 2π en $\theta(t)$; *a*) y *b*) son gráficas de $\theta(t)$ y $\theta'(t)$, respectivamente.

A partir del diagrama fasorial de la figura 2.43, es posible deducir las condiciones que se requieren para que ocurran los clics. Un clic que se hace positivo ocurre cuando la envolvente $r(t)$ y la fase $\psi(t)$ del ruido de banda angosta $n(t)$ satisfacen las siguientes condiciones:

$$r(t) > A_c$$
$$\psi(t) < \pi \leq \psi(t) + d\,\psi(t)$$
$$\frac{d\,\psi(t)}{dt} > 0$$

Estas condiciones aseguran que la fase $\theta(t)$ del fasor resultante $x(t)$ cambia en 2π radianes en el incremento de tiempo dt, durante el cual la fase del ruido de banda angosta aumenta en la cantidad incremental $d\psi(t)$. De modo similar, las condiciones para que ocurra un clic que se vuelve negativo son las siguientes:

$$r(t) > A_c$$
$$\psi(t) > -\pi > \psi(t) + d\,\psi(t)$$
$$\frac{d\,\psi(t)}{dt} < 0$$

Estas condiciones aseguran que $\theta(t)$ cambia en -2π radianes durante el incremento de tiempo dt.

La *relación de portadora a ruido* se define por medio de

$$\rho = \frac{A_c^2}{2B_T N_0} \tag{2.154}$$

Cuando ρ disminuye, aumenta el número promedio de clics por unidad de tiempo. Cuando este número se vuelve apreciablemente grande, se dice que se presenta el umbral.

La relación señal a ruido de salida se calcula como sigue:

1. La señal de salida se toma cuando la salida del receptor se mide sin ruido. La potencia promedio de la señal de salida se calcula suponiendo una modulación senoidal que produce una desviación de frecuencia Δf igual a $B_T/2$, por lo que la portadora oscila a través de la banda completa de la frecuencia de entrada.
2. La potencia promedio del ruido de salida se calcula cuando no hay señal presente; esto es, la portadora no está modulada, sin que se imponga restricción en el valor de la relación de portadora a ruido ρ.

De acuerdo con esta base heurística, la teoría[11] produce la curva I de la figura 2.45 que presenta una gráfica de la relación señal a ruido de salida en función de la relación portadora al ruido cuando la proporción $B_T/2W$ es igual a 5. Esta curva muestra que la relación señal a ruido de salida se aparta en forma apreciable de una función lineal de la relación portadora a ruido ρ, cuando ρ es menor que aproximadamente 10 dB. La curva II de la figura 2.45 muestra el efecto de la modulación en la relación señal a ruido de salida cuando la señal moduladora (supuesta senoidal) y el ruido están presentes al mismo tiempo. La potencia promedio de la señal de salida perteneciente a la curva II puede considerarse efectivamente la misma que la correspondiente a la curva I. Sin embargo, la potencia promedio del ruido de salida depende en forma considerable de la presencia de la señal moduladora, lo cual explica la notable desviación de la curva II con respecto a la curva I. En particular, encontramos que cuando ρ disminuye desde el infinito, la señal de salida a ruido se desvía apreciablemente de la función lineal de ρ, cuando esta última corresponde aproximada-

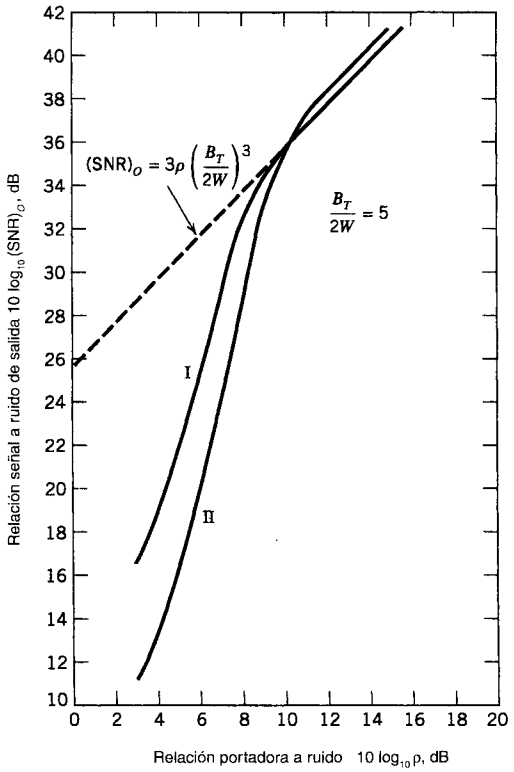


FIGURA 2.45 Dependencia de la relación señal a ruido de salida con respecto a la relación de portadora de entrada a ruido para un receptor de FM. En la curva I, la potencia promedio del ruido de salida se calcula suponiendo una portadora modulada senoidalmente. En la curva II, la potencia promedio del ruido de salida se calcula suponiendo una portadora modulada senoidalmente. Ambas curvas se calculan con base en la teoría.

mente a 11 dB. Asimismo, cuando la señal está presente, la modulación resultante de la portadora tiende a aumentar el número promedio de clics por segundo. Experimentalmente, se encuentra que se escuchan clics ocasionales en la salida del receptor a una relación portadora a ruido cercana a 13 dB, la cual parece ser sólo ligeramente más alta que la que indica la teoría. Resulta también de interés notar que el aumento en el número promedio de clics por segundo tiende a provocar que la relación señal a ruido de salida descienda en forma un poco más pronunciada justo por debajo del nivel de umbral en la presencia de modulación.

De acuerdo con la exposición anterior, es posible concluir que los efectos de umbral en los receptores de FM pueden evitarse en la mayor parte de los casos prácticos de interés si la relación portadora a ruido ρ es igual o mayor que 20 o, equivalentemente, 13 dB. De ese modo, al utilizar la ecuación (2.154), encontramos que la pérdida de mensaje en la salida del discriminador es despreciable si

$$\frac{A_c^2}{2B_T N_0} \geq 20$$

o, de modo equivalente, si la potencia transmitida promedio $A_c^2/2$ satisface la condición

$$\frac{A_c^2}{2} \geq 20B_T N_0 \tag{2.155}$$

Para utilizar esta fórmula, es posible que procedamos de la manera siguiente:

1. Para un índice de modulación β y un ancho de banda del mensaje W especificados, determinamos el ancho de banda de transmisión de la onda de FM, B_T, utilizando la curva universal de la figura 2.26 o la regla de Carson.
2. Para una potencia de ruido promedio especificada por ancho de banda unitaria, N_0, usamos la ecuación (2.155) para determinar el valor mínimo de la potencia transmitida promedio $A_c^2/2$ que se necesita para operar arriba del umbral.

■ REDUCCIÓN DEL UMBRAL DE FM

En los sistemas de comunicación que utilizan modulación de frecuencia, existe un interés particular por reducir el umbral de ruido en un receptor de FM de modo que opere satisfactoriamente el receptor dentro de la mínima potencia posible de la señal. La reducción de umbral en los receptores de FM puede conseguirse utilizando un demodulador de FM con realimentación negativa[12] (conocido por lo común como *demodulador* FMFB) o utilizando un *demodulador de lazo de cierre de fase*. Estos dispositivos se conocen como *demoduladores de umbral extendido* (la idea de los mismos se ilustra en la figura 2.46). La extensión del umbral que se muestra en esta figura se mide en relación con el discriminador de frecuencia estándar (es decir, uno sin realimentación).

En la figura 2.47 se presenta un diagrama de bloques de un demodulador FMFB.[13] Observamos que el oscilador local del receptor convencional de FM se ha sustituido por un oscilador controlado por voltaje (VCO) cuya frecuencia de salida instantánea se controla mediante una señal demodulada. Para entender la operación de este receptor, suponga por el momento que el oscilador controlado por voltaje se elimina del circuito y que el lazo de realimentación se deja abierto. Considere que se aplica una señal de FM de banda ancha a la entrada del receptor y que una segunda señal de FM, proveniente de la misma fuente pero cuyo índice de modulación es una fracción más pequeña, se aplica a la terminal del VCO del mezclador. La salida del mezclador consistiría en la componente de frecuencia de diferencia, ya que la componente de frecuencia de la suma la elimina el filtro pasobanda. La desviación de frecuencia de la salida del mezclador sería pequeña, aunque la desviación de frecuencia de ambas ondas de FM de entrada es grande, puesto que es pequeña la diferencia entre sus desviaciones instantáneas. Por tanto, los índices de modulación se restarían y la onda de FM resultante en la salida del mezclador

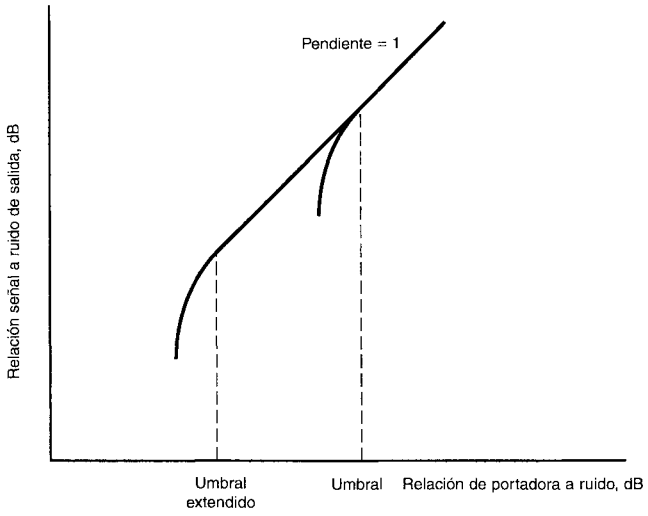


FIGURA 2.46 Extensión del umbral de FM.

tendría un índice de modulación más pequeño. La onda de FM con índice de modulación reducido puede hacerse pasar por un filtro pasobanda, cuyo ancho de banda sólo necesita ser una fracción de la que se requiere para cualquiera de las FM de banda ancha, y luego demodularse la frecuencia. Es claro que en estas condiciones puede obtenerse la segunda señal de FM de banda ancha aplicada al mezclador realimentando la salida del discriminador de frecuencia nuevamente al oscilador controlado por voltaje.

Se demostrará ahora que la relación de señal a ruido de un receptor FMFB es la misma que la de un receptor de FM convencional con la misma señal de entrada y potencia de ruido si la relación portadora a ruido es suficientemente grande. Suponga por el momento que no hay realimentación alrededor del demodulador. En la presencia combinada de una portadora no modulada $A_c \cos(2\pi f_c t)$ y ruido de banda angosta

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t)$$

la fase de la señal compuesta $x(t)$ en la entrada del limitador-discriminador es aproximadamente igual a $n_Q(t)/A_c$, suponiendo que la relación portadora a ruido es alta. La envolvente de $x(t)$ no nos interesa, ya que el limitador elimina todas las variaciones en la envolvente. De ese modo, la señal compuesta en la entrada del discriminador de frecuencia consiste en una onda modulada en fase de pequeño índice con la modulación obtenida de la componente $n_Q(t)$ del ruido que está en cuadratura de fase con la portadora. Cuando se aplica la realimentación, el oscilador controlado por voltaje genera una señal modulada en frecuencia que reduce el índice de modulación en fase de la onda a la salida del filtro pasobanda, esto es, la componente en cuadratura $n_Q(t)$ del ruido. Vemos de ese modo que siempre que la relación de portadora a ruido es lo suficientemente grande, el receptor de FMFB no responde a la componente de ruido en fase $n_I(t)$, aunque demodularía la componente de ruido en cuadratura $n_Q(t)$ exactamente del mismo modo que demodularía la modulación de la señal. La señal

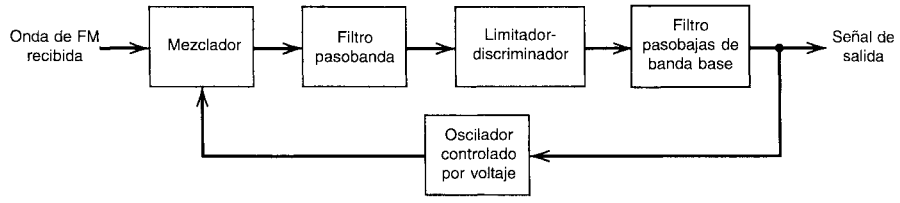


FIGURA 2.47 Demodulador de FM con realimentación negativa.

y el ruido en cuadratura se reducen en la misma proporción mediante la realimentación aplicada, resultando que la relación de la señal a ruido de banda base es independiente de la realimentación. Para altas relaciones de portadora a ruido, la relación de la señal a ruido de banda base de un receptor de FMFB es entonces la misma que la de un receptor de FM convencional.

La razón de que un receptor de FMFB sea capaz de extender el umbral es que, a diferencia del receptor de FM convencional, utiliza una pieza importante de información *a priori*, a saber, que aun cuando la frecuencia de la portadora de la onda de FM entrante por lo común tendrá grandes desviaciones de frecuencia, su velocidad de cambio estará en la velocidad de la banda base. Un demodulador de FMFB es en esencia un *filtro de rastreo* que puede seguir únicamente la frecuencia que varía en forma lenta de una señal de FM de banda ancha y, en consecuencia, responde sólo a una banda angosta de ruido centrado en torno a la frecuencia instantánea de la portadora. El ancho de banda del ruido al cual responde el receptor de FMFB es precisamente la banda de ruido que rastrea el oscilador controlado por voltaje. El resultado final radica en que un receptor de FMFB es capaz de realizar una extensión de umbral del orden de 5–7 dB, lo cual representa una mejora importante en el diseño de sistemas de FM de potencia mínima.

Al igual que el demodulador de FMFB, la malla de fase encadenada (que se explica más adelante en la sección 2.14) también es un filtro de rastreo y, como tal, el ancho de banda de ruido al que responde es precisamente la banda de ruido que rastrea el oscilador controlado por voltaje. En realidad, el demodulador de malla de fase encadenada ofrece una capacidad de extensión de umbral con un circuito más o menos simple. Desafortunadamente, la cantidad de extensión de umbral no la predice ninguna teoría disponible y depende de los parámetros de la señal. En general, se consigue una mejora de unos cuantos decibeles (del orden de 2 a 3) en las aplicaciones características, lo que no es tan bueno como un demodulador de FMFB.

■ PREÉNFASIS Y DEÉNFASIS EN FM

La ecuación (2.147) demuestra que la densidad espectral de potencia del ruido a la salida de un receptor de FM tiene una dependencia de ley cuadrática sobre la frecuencia de operación. Esto se ilustra en la figura 2.48*a*. En la figura 2.48*b* se incluye la densidad espectral de potencia de una fuente de mensaje característica; las señales de audio y video tienen por lo general espectros de esta forma. En particular, la densidad espectral de potencia del mensaje casi siempre desciende en forma apreciable a frecuencias superiores. Por otra parte, la densidad espectral de potencia del ruido de salida aumenta con rapidez con la frecuencia. Así, alrededor de $f = \pm W$, la densidad espectral relativa del mensaje es bastante baja, en tanto que la del ruido de salida es comparativamente bastante elevada. Sin duda, el mensaje no está usando de manera eficiente la banda de frecuencia que se le asigna. Es posible pensar que una forma de mejorar el desempeño del sistema ante el ruido consista en reducir un poco el ancho de banda del filtro pasobajas posdetección, de modo que rechace una gran cantidad de potencia de ruido mientras se pierde únicamente una pequeña cantidad de la potencia del mensaje. Sin embargo, un método de este tipo suele ser insatisfactorio porque la distorsión del mensaje que provoca el ancho de banda reducido del filtro, si bien ligero, quizá no sea tolerable. Por ejemplo, en el caso de la música se observa que a pesar de que las notas de alta frecuencia aportan sólo una muy pequeña fracción de la potencia total, contribuyen en gran medida desde el punto de vista estético.

Un método más satisfactorio para el uso eficiente de la banda de frecuencia asignada se basa en el uso de preénfasis en el transmisor y deénfasis en el receptor, como se ilustra en la

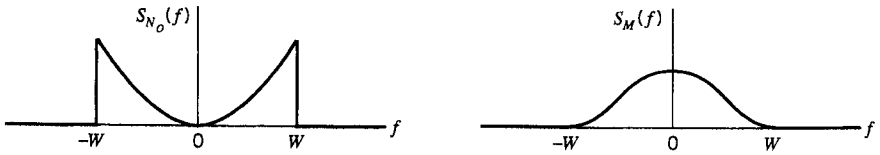


FIGURA 2.48 *a*) Densidad espectral de potencia del ruido a la salida de receptor de salida. *b*) Densidad espectral de potencia de una señal de mensaje característica.

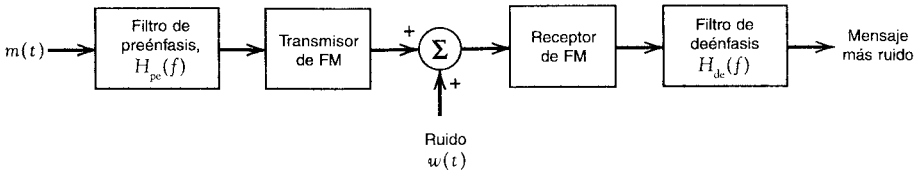


FIGURA 2.49 Empleo de preénfasis y deénfasis en un sistema de FM.

figura 2.49. En este método se preenfatizan en forma artificial las componentes de alta frecuencia de la señal del mensaje antes de la modulación en el transmisor y, por tanto, antes de que se introduzca el ruido en el receptor. En efecto, las porciones de baja y alta frecuencia de la densidad espectral de potencia del mensaje se ecualizan de manera tal que el mensaje ocupa por completo la banda de frecuencia que se le asigna. Entonces, a la salida del discriminador en el receptor realizamos la operación inversa deénfasis en las componentes de alta frecuencia, de modo que se restaure la distribución original de la potencia de señal del mensaje. En un proceso de este tipo, también se reducen las componentes de alta frecuencia del ruido a la salida del discriminador, aumentándose por ello en forma efectiva la relación señal a ruido de salida del sistema. Tales procesos de preénfasis y deénfasis se usan ampliamente en la transmisión y la recepción comercial de radio FM.

Para producir una versión no distorsionada del mensaje original en la salida del receptor, el filtro de preénfasis en el transmisor y el filtro de deénfasis en el receptor deben tener idealmente respuestas en frecuencia que son la inversa una de otra. Esto es, si $H_{pe}(f)$ designa la respuesta en frecuencia del filtro de preénfasis, entonces la respuesta en frecuencia $H_{de}(f)$ del filtro de deénfasis debe ser idealmente (ignorando el retardo de la transmisión)

$$H_{de}(f) = \frac{1}{H_{pe}(f)}, \quad -W \le f \le W \tag{2.156}$$

Esta elección de la respuesta en frecuencia hace que la potencia promedio del mensaje en la salida del receptor sea independiente del proceso de preénfasis y deénfasis.

A partir de nuestro análisis de ruido anterior en sistemas de FM, suponiendo una relación portadora a ruido elevada, la densidad espectral de potencia del ruido $n_d(t)$ a la salida del discriminador está dada por la ecuación (2.146). La densidad espectral de potencia modificada del ruido a la salida del filtro de deénfasis es entonces

$$|H_{de}(f)|^2 S_{N_d}(f) = \begin{cases} \dfrac{N_0 f^2}{A_c^2}|H_{de}(f)|^2, & |f| \le \dfrac{B_T}{2} \\ 0, & \text{en otro caso} \end{cases} \tag{2.157}$$

Admitiendo, como antes, que el filtro pasobajas de posdetección tiene un ancho de banda W que es, en general, menor que $B_T/2$, encontramos que la potencia promedio del ruido modificado a la salida del receptor es como sigue:

$$\begin{pmatrix} \text{Potencia promedio del} \\ \text{ruido de salida con deénfasis} \end{pmatrix} = \frac{N_0}{A_c^2}\int_{-W}^{W} f^2 |H_{de}(f)|^2 \, df \tag{2.158}$$

Debido a que la potencia promedio del mensaje en la salida del receptor idealmente no se ve afectada por el procedimiento combinado de preénfasis y deénfasis, se concluye que la mejora en la relación señal a ruido de salida que se produce mediante preénfasis en el transmisor y deénfasis en el receptor está definida por

$$I = \frac{\text{potencia promedio del ruido de salida sin preénfasis y deénfasis}}{\text{potencia promedio del ruido de salida con preénfasis y deénfasis}} \tag{2.159}$$

Demostramos antes que la potencia promedio del ruido de salida sin preénfasis y deénfasis es igual a $(2N_0W^3/3A_c^2)$; vea la ecuación (2.148). Por tanto, después de la cancelación de los términos comunes, podemos expresar el factor de mejoramiento I, como

$$I = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2 \, df} \tag{2.160}$$

Debe subrayarse que este factor de mejora supone el uso de una relación de portadora a ruido alta a la entrada del discriminador en el receptor.

EJEMPLO 2.6

Un filtro de preénfasis simple que enfatiza las frecuencias elevadas y que se usa comúnmente en la práctica se define por medio de la respuesta en frecuencia

$$H_{pe}(f) = 1 + \frac{jf}{f_0}$$

la cual se ejecuta bastante bien mediante la red de amplificador RC que se muestra en la figura 2.50*a*, siempre que $R << r$ y $2\pi fCr << 1$ dentro de la banda de frecuencias de interés. El amplificador en la figura 2.50*a* persigue compensar la atenuación que introduce la red RC a bajas frecuencias. El parámetro de frecuencia f_0 es $1/(2\pi Cr)$.

El filtro de deénfasis correspondiente en el receptor se define por medio de la respuesta en frecuencia

$$H_{de}(f) = \frac{1}{1 + jf/f_0}$$

que puede implantarse utilizando la red RC simple de la figura 2.50*b*.

La mejora en la relación de la señal de salida al ruido del receptor de FM, que resulta del uso combinado de los filtros de preénfasis y deénfasis de la figura 2.50, es consecuentemente

$$\begin{aligned} I &= \frac{2W^3}{3\int_{-W}^{W} \frac{f^2 \, df}{1 + (f/f_0)^2}} \\ &= \frac{(W/f_0)^3}{3[(W/f_0) - \tan^{-1}(W/f_0)]} \end{aligned} \tag{2.161}$$

En la transmisión comercial de FM es característico tener $f_0 = 2.1$ kHz, y es posible suponer razonablemente que $W = 15$ kHz. Este conjunto de valores produce $I = 22$, que corresponde a una mejora de 13 dB en la relación señal de salida a ruido de receptor. Esta relación para un receptor de FM sin preénfasis y deénfasis corresponde por lo común a valores de 40—50 dB. Observamos, por lo

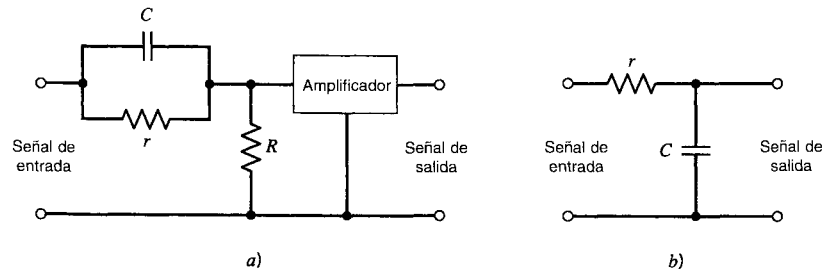


FIGURA 2.50 *a*) Filtro de preénfasis. *b*) Filtro de deénfasis.

tanto, que al utilizar los filtros simples de preénfasis y deénfasis de la figura 2.50, podemos poner en práctica una mejora importante en el desempeño del receptor ante el ruido. ◀

El uso de los filtros *lineales* simples de preénfasis y deénfasis que acaban de describirse es un ejemplo de cómo el desempeño de un sistema de FM puede mejorarse utilizando las diferencias entre las características de las señales y el ruido del sistema. Estos filtros simples también se aplican en la grabación de cintas de audio. Específicamente, se han aplicado con éxito técnicas *no lineales* de preenfatización y deenfatización en la grabación de cintas. Estas técnicas[14] (conocidas como sistema *Dolby-A*, *Dolby-B* y *DBX*) usan una combinación de filtrado y de compresión del intervalo dinámico para reducir los efectos del ruido, particularmente cuando es bajo el nivel de la señal.

2.14 *Experimentos de computadora: malla de fase encadenada*

El estudio experimental que se presenta en esta sección se enfoca en el uso de la malla de fase encadenada en la demodulación de una señal modulada en frecuencia. Sin embargo, antes de avanzar con los experimentos presentamos una breve exposición de la teoría de la malla de fase encadenada.

Básicamente *la malla de fase encadenada* consta de tres componentes principales: un *multiplicador*, un *filtro de lazo* y un *oscilador controlado por voltaje* (VCO) conectados entre sí en la forma de un sistema realimentado, como se indica en la figura 2.51. El oscilador controlado por voltaje es un generador senoidal cuya frecuencia se determina por medio de un voltaje aplicado a él desde una fuente externa. En efecto, cualquier modulador de frecuencia puede servir como un oscilador de este tipo. Suponemos que al principio se ha ajustado el controlador de manera que cuando el voltaje de control es cero, se satisfacen dos condiciones:

1. La frecuencia del oscilador se fija precisamente a la frecuencia de portadora no modulada f_c.
2. La salida del oscilador controlado por voltaje tiene un corrimiento de fase de 90 grados con respecto a la onda portadora no modulada.

Suponga entonces que la señal de entrada aplicada a la malla de fase encadenada es una señal de FM definida por

$$s(t) = A_c \operatorname{sen}[2\pi f_c t + \phi_1(t)]$$

donde A_c es la amplitud de la portadora. Con una señal moduladora $m(t)$, el ángulo $\phi_1(t)$ se relaciona con $m(t)$ mediante la integral

$$\phi_1(t) = 2\pi k_f \int_0^t m(\tau)\, d\tau$$

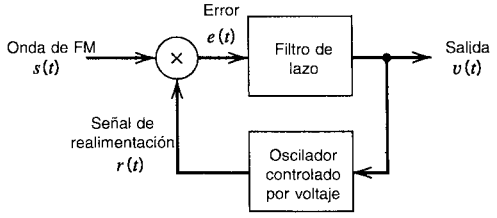


FIGURA 2.51 Malla de fase encadenada.

donde k_f es la sensibilidad de frecuencia del modulador de frecuencia. Considérese que la salida del oscilador controlado por voltaje en la malla de fase encadenada está definida por

$$r(t) = A_v \cos[2\pi f_c t + \phi_2(t)]$$

donde A_v es la amplitud. Con el voltaje de control $v(t)$ aplicado a la entrada del oscilador, el ángulo $\phi_2(t)$ se relaciona con $v(t)$ mediante la integral

$$\phi_2(t) = 2\pi k_v \int_0^t v(\tau)\, d\tau \tag{2.162}$$

donde k_v es la sensibilidad de frecuencia del oscilador, medida en hertz por volt. El objetivo de la malla de fase encadenada es generar una salida del oscilador $r(t)$ que tiene el mismo ángulo de fase (salvo por la diferencia fija de 90 grados) que la señal de entrada $s(t)$ de FM. El ángulo de fase variable en el tiempo $\phi_1(t)$ que caracteriza a $s(t)$ tal vez se deba a la modulación mediante una señal de mensaje $m(t)$, en cuyo caso deseamos recuperar $\phi_1(t)$ y de ese modo producir una estimación de $m(t)$. En otras aplicaciones de la malla de fase encadenada, el ángulo de fase variable en el tiempo $\phi_1(t)$ de la señal entrante $s(t)$ puede ser un cierre de fase indeseable provocado por fluctuaciones en el canal de comunicación; en este último caso, deseamos *rastrear* $\phi_1(t)$ de modo que se produzca una señal con el mismo ángulo de fase con fines de detección coherente (de modulación síncrona).

■ MODELO DE LA MALLA DE FASE ENCADENADA[15]

Para entender la malla de fase encadenada, es deseable tener un *modelo* del lazo. Empezamos formulando un modelo no lineal, que subsecuentemente se hace lineal para simplificar el análisis. De acuerdo con la figura 2.51, la señal de FM entrante $s(t)$ y la salida $r(t)$ del oscilador controlado por voltaje se aplican al multiplicador, produciéndose dos componentes:

1. Una componente de alta frecuencia, representada por el término de *doble frecuencia*

$$k_m A_c A_v \operatorname{sen}[4\pi f_c t + \phi_1(t) + \phi_2(t)]$$

2. Una componente de baja frecuencia, representada por el término de *diferencia de frecuencia*

$$k_m A_c A_v \operatorname{sen}[\phi_1(t) - \phi_2(t)]$$

donde k_m es la *ganancia del multiplicador*, medida en volts^{-1}.

El filtro de lazo en malla de fase encadenada es uno pasobajas, y su respuesta a la componente de alta frecuencia puede ignorarse. El oscilador controlado por voltaje contribuye también a la atenuación de esta componente. Por tanto, descartando la componente de alta frecuencia (es decir, el término de doble frecuencia), la entrada al filtro del lazo se reduce a

$$e(t) = k_m A_c A_v \operatorname{sen}[\phi_e(t)] \tag{2.163}$$

donde $\phi_e(t)$ es el *error de fase* definido por

$$\begin{aligned} \phi_e(t) &= \phi_1(t) - \phi_2(t) \\ &= \phi_1(t) - 2\pi k_v \int_0^t v(\tau)\, d\tau \end{aligned} \tag{2.164}$$

El filtro de lazo opera en el error $e(t)$ para producir una salida $v(t)$ definida por la integral de convolución:

$$v(t) = \int_{-\infty}^{\infty} e(\tau) h(t - \tau)\, d\tau \tag{2.165}$$

donde $h(t)$ es la respuesta al impulso del filtro de lazo. Utilizando las ecuaciones (2.164) y (2.165) para relacionar $\phi_e(t)$ y $\phi_1(t)$, obtenemos la siguiente ecuación íntegro-diferencial no lineal como el descriptor del comportamiento dinámico de la malla de fase encadenada:

$$\frac{d\phi_e(t)}{dt} = \frac{d\phi_1(t)}{dt} - 2\pi K_0 \int_{-\infty}^{\infty} \text{sen}[\phi_e(\tau)] h(t-\tau)\, d\tau \tag{2.166}$$

donde K_0 es un *parámetro de ganancia de lazo* definido por

$$K_0 = k_m k_v A_c A_v \tag{2.167}$$

Las amplitudes A_c y A_v se miden ambas en volts, la ganancia del multiplicador k_m, en volt^{-1}, y la sensibilidad de frecuencia k_v, en hertz por volt. Por tanto, se concluye de la ecuación (2.167) que K_0 tiene las dimensiones de frecuencia. La ecuación (2.166) sugiere el modelo que se muestra en la figura 2.52 para una malla de fase encadenada. En él hemos incluido también la relación entre $v(t)$ y $e(t)$ según la representan las ecuaciones (2.163) y (2.165). Vemos que el modelo de la figura 2.52 se asemeja al diagrama de bloques real de la figura 2.51. El multiplicador a la entrada de la malla de fase encadenada se sustituye por un restador y una no linealidad senoidal, y el oscilador controlado por voltaje por un integrador.

La no linealidad senoidal en el modelo de la figura 2.52 complica la tarea de analizar el comportamiento de la malla de fase encadenada. Resultaría útil *linealizar* este modelo para simplificar el análisis y obtener aun así una buena descripción aproximada del comportamiento del lazo en ciertos modos de operación. Cuando el error de fase $\phi_e(t)$ es cero, se dice que la malla de fase encadenada estará en *enganche*. Cuando $\phi_e(t)$ es pequeña en todo momento en comparación con un radián, es posible utilizar la aproximación

$$\text{sen}[\phi_e(t)] \simeq \phi_e(t)$$

que es exacta hasta dentro de 4% para $\phi_e(t)$ menor que 0.5 radianes. En este caso, se dice que el lazo estará *casi en enganche*, y la no linealidad senoidal de la figura 2.52 puede descartarse. Bajo esta condición, $v(t)$ es aproximadamente igual a $m(t)$, salvo por el factor de escalamiento k_f/k_v.

La complejidad de la malla de fase encadenada está determinada por la respuesta en frecuencia $H(f)$ del filtro. La forma más simple de una malla de fase encadenada se obtiene cuando $H(f) = 1$, esto es, no hay filtro de lazo, y el que resulta se conoce como *malla de fase encadenada de primer orden*. Una limitación principal de este último tipo de lazo es que el parámetro de ganancia de lazo K_0 controla tanto el ancho de banda de lazo como el intervalo *de* frecuencia de mantenimiento del mismo; el *intervalo de frecuencia de mantenimiento* se refiere a la gama de frecuencias para las cuales el

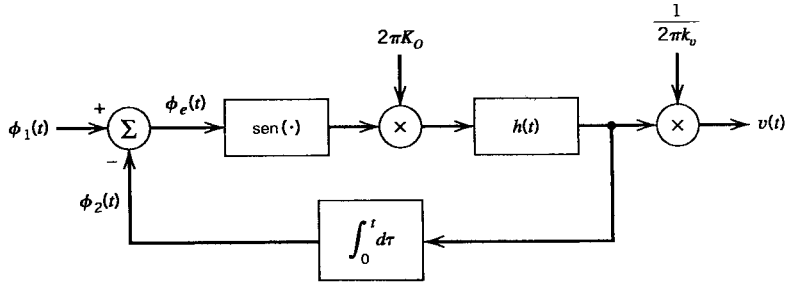


FIGURA 2.52 Modelo no lineal de la malla de fase encadenada.

lazo permanece enganchado ante la señal de entrada. Podemos superar esta limitación utilizando un filtro de lazo con la respuesta en frecuencia

$$H(f) = 1 + \frac{a}{jf} \tag{2.168}$$

donde a es una constante. Con la aplicación de este filtro de lazo y la operación de la malla de fase encadenada en su modo lineal, encontramos de acuerdo con la ecuación (2.166) que la malla de fase encadenada se comporta como un *sistema realimentado de segundo orden*, como muestra la respuesta en frecuencia estándar

$$\frac{\Phi_e(f)}{\Phi_1(f)} = \frac{(jf/f_n)^2}{1 + 2\zeta(jf/f_n) + (jf/f_n)^2} \tag{2.169}$$

donde $\Phi_e(f)$ y $\Phi_1(f)$ son las transformadas de Fourier de $\phi_e(t)$ y $\phi_1(t)$, respectivamente. El sistema está parametrizado por la *frecuencia natural*, f_n, y por el *factor de amortiguamiento*, ζ, que se definen respectivamente por medio de

$$f_n = \sqrt{aK_0} \tag{2.170}$$

y

$$\zeta = \sqrt{\frac{K_0}{4a}} \tag{2.171}$$

La malla de fase encadenada de segundo orden descrita de esta manera es el tema de los experimentos de computadora que se presentan a continuación.

Experimento 1: modo de adquisición

Cuando una malla de fase encadenada se usa para detección coherente (demodulación síncrona), el lazo se debe cerrar primero sobre la señal de entrada y luego seguir las variaciones de su ángulo de fase con el tiempo. El proceso de llevar una malla al enganche se denomina *adquisición* y el proceso consecuente de seguir las variaciones angulares en la señal de entrada se conoce como *rastreo*. En el modo de adquisición y muy posiblemente en el modo de rastreo, el error de $\phi_e(t)$ entre la señal de entrada $s(t)$ y la salida $r(t)$ del oscilador controlado por voltaje será indudablemente grande, demandando por ello el uso del modelo no lineal de la figura 2.52. Sin embargo, un análisis no lineal del proceso de adquisición basado en este último modelo está más allá de los fines de este libro. En este experimento recurrimos a simulaciones de computadora para estudiar el proceso de adquisición y de ese modo obtener un conocimiento amplio de algunas de sus características.

Considere una malla de fase encadenada de segundo orden utilizando el filtro de lazo de la ecuación (2.168) y los siguientes parámetros:

$$\text{Frecuencia natural } f_n = \frac{1}{2\pi}\text{ Hz}$$

$$\text{Factor de amortiguamiento } \zeta = 0.3, 0.707, 1.0$$

Para aceptar la variación en ζ, el parámetro a del filtro se modifica de acuerdo con la fórmula

$$a = \frac{f_n}{2\zeta}$$

que se desprende de las ecuaciones (2.170) y (2.171). La figura 2.53 presenta la variación en el error de fase $\phi_e(t)$ con el tiempo para cada uno de los tres valores especificados del factor de

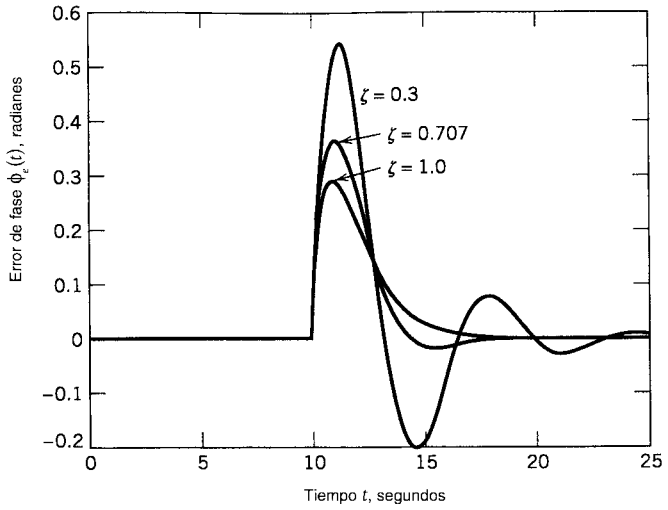


FIGURA 2.53 Variación del error de fase para tres valores diferentes del factor de amortiguamiento.

amortiguamiento ζ, suponiendo un paso de frecuencia de 0.125 Hz. Estos resultados muestran que el factor de amortiguamiento $\zeta = 0.707$ produce el mejor compromiso entre el tiempo de respuesta rápida y el comportamiento oscilatorio subamortiguado.

Experimento 2: descripción gráfica en el plano de fase

Una *descripción gráfica en el plano de fase* es una familia de trayectorias, con cada una representando una solución única de la ecuación (2.166). En el segundo experimento se muestra la descripción gráfica en el plano de fase de una malla de fase encadenada de segundo orden para el caso de una modulación senoidal. Los parámetros del sistema de lazo son como sigue:

$$\text{Parámetro de la ganancia del lazo } K_0 = \frac{50}{2\pi}\text{Hz}$$

$$\text{Frecuencia natural del lazo } f_n = \frac{50}{2\sqrt{2}}\text{Hz}$$

$$\text{Frecuencia de modulación senoidal } f_m = \frac{50}{2\pi\sqrt{2\pi}}\text{Hz}$$

La figura 2.54 presenta la descripción gráfica en el plano de fase de la malla de fase encadenada ajustado para amortiguamiento crítico, donde las trayectorias (error de la frecuencia en función del error de fase) se grafican para diferentes puntos de inicio. A partir de esta descripción gráfica hacemos las siguientes observaciones:

1. En una no linealidad senoidal, la descripción gráfica en el plano de fase es en sí misma periódica con periodo 2π en el error de fase ϕ_e, aunque es aperiódica en $d\phi_e/dt$.
2. Para un error de frecuencia inicial

$$\frac{1}{K}\frac{d\phi_e}{dt}$$

con un valor absoluto menor que o igual a 1, se asegura que la malla de fase encadenada alcance un *punto estable* (*equilibrio*) en (0, 0) o $(0, 2\pi)$; la multiplicidad de puntos de equilibrio es una manifestación de la periodicidad de la descripción gráfica en el plano de fase.

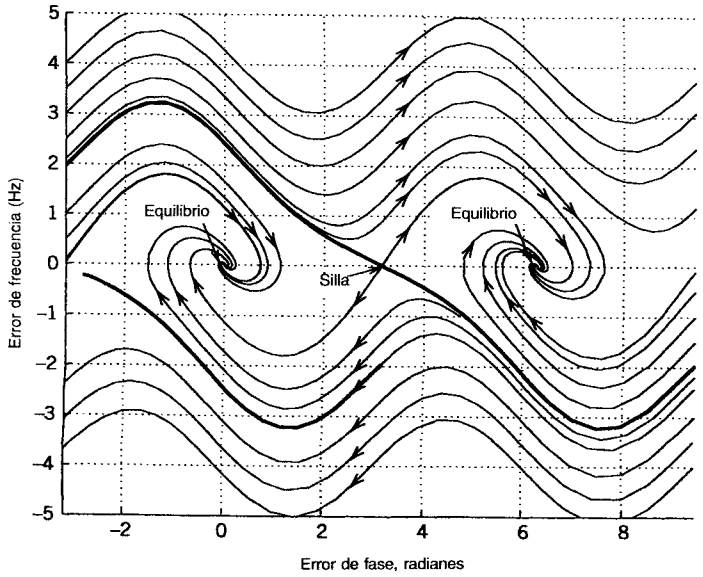


FIGURA 2.54 Descripción gráfica en el plano de fase para amortiguamiento crítico y modulación senoidal.

3. Para un error de frecuencia inicial

$$\frac{1}{K}\frac{d\phi_e}{dt}$$

con un valor absoluto igual a 2, tenemos un *punto de silla* en $(0,\pi)$ en el que la perturbación más ligera aplicada a la malla de fase encadenada ocasiona un corrimiento del punto de equilibrio (0, 0) o $(0, 2\pi)$.

2.15 *Resumen y discusión*

En este capítulo estudiamos los principios de la modulación de onda continua. Esta forma analógica de modulación utiliza una portadora senoidal cuya amplitud o ángulo se varía de acuerdo con la señal del mensaje. De ese modo podemos distinguir dos familias de modulación de onda continua: modulación de amplitud y modulación angular.

■ MODULACIÓN DE AMPLITUD

La modulación de amplitud puede clasificarse en cuatro tipos, dependiendo del contenido espectral de la señal modulada. Los cuatros tipos de modulación de amplitud y sus méritos prácticos son los siguientes:

1. *Modulación de amplitud completa (AM), en la cual las bandas laterales superior e inferior se transmiten completas, acompañadas por la onda portadora.*
En consecuencia, la demodulación de una señal de AM se lleva a cabo de manera muy simple en el receptor utilizando un detector de envolvente, por ejemplo. Por esta razón encontramos que la AM completa suele usarse en la *transmisión de radio* de AM comercial, la cual implica un solo y poderoso transmisor y numerosos receptores cuya construcción es relativamente económica.

2. *Modulación de doble banda lateral–portadora suprimida (DSB-SC), en la cual sólo se transmiten las bandas laterales superior e inferior.*
La supresión de la onda portadora significa que la modulación de DSB-SC requiere mucho menor potencia que la AM completa para transmitir la misma señal de mensaje; sin embargo, esta ventaja de la modulación de DSB-SC sobre la AM completa se alcanza a costa de una mayor complejidad del receptor. Por tanto, la modulación de DSB-SC resulta adecuada para la *comunicación punto a punto* que implica un transmisor y un receptor; en esta forma de comunicación, la potencia transmitida es muy solicitada y en ese caso se justifica el uso de un receptor complejo.

3. *Modulación de banda lateral única (SSB), en la cual sólo se transmite la banda lateral superior o la inferior.*
La modulación de SSB es la forma óptima de la modulación de onda continua en el sentido de que requiere la mínima potencia transmitida y el mínimo ancho de banda de canal para transportar una señal de mensaje de un punto a otro. Sin embargo, su uso se limita a señales de mensaje con una brecha de energía centrada en la frecuencia cero.

4. *Modulación de banda lateral residual (VSB), en la que casi la totalidad de una banda lateral y un residuo de la otra banda lateral se transmiten de un modo complementario y preestablecido.*
La modulación de VSB requiere un ancho de banda de canal que se encuentra entre el que se requiere para los sistemas de SSB y de DSB-SC, y el ahorro en ancho de banda puede ser importante si se manejan señales moduladoras con anchos de banda grandes, como en el caso de las señales de televisión y los datos de alta velocidad.

La DSB-SC, la SSB y la VSB constituyen ejemplos de modulación lineal, en tanto que, en términos generales, la AM completa es no lineal. Sin embargo, la desviación con respecto a la linealidad de la AM completa es de tipo medio. En consecuencia, las cuatro formas de modulación de amplitud llevan de inmediato por sí solas al análisis espectral utilizando la transformada de Fourier.

■ Modulación angular

La modulación angular puede clasificarse como modulación en frecuencia (FM) y modulación de fase (PM). *En la FM, la frecuencia instantánea de una portadora senoidal se varía en proporción a la señal de mensaje. En la PM, en cambio, la fase portadora es la que se varía en proporción a la señal de mensaje.* La frecuencia instantánea se define como la derivada de la fase con respecto al tiempo, salvo por el factor de escalamiento $1/(2\pi)$. Por tanto, la FM y la PM guardan entre sí una estrecha relación; si conocemos las propiedades de una, es posible determinar las de la otra. Por esta razón, y debido a que la FM suele usarse en la transmisión, buena parte del material acerca de la modulación angular en el capítulo se dedicó a la modulación de frecuencia.

A diferencia de la modulación de amplitud, la FM es un proceso de modulación no lineal. En consecuencia, el análisis espectral de la FM es más difícil que el de la AM. A pesar de eso, al estudiar la FM de un solo tono, fuimos capaces de crear un conocimiento amplio de las propiedades espectrales de la FM. Obtuvimos en particular una regla empírica conocida como regla de Carson para una evaluación aproximada del ancho de banda de transmisión B_T de la FM. De acuerdo con esta regla, B_T está controlada por un solo parámetro: el índice de modulación β para la FM senoidal o la razón de desviación D para la FM no senoidal.

■ Análisis de ruido

Concluimos el capítulo acerca de los sistemas de modulación de onda continua presentando una comparación de sus desempeños ante el ruido. Supusimos en ello que la modulación se produce mediante una onda senoidal. Para que la comparación fuera útil, consideramos también que los

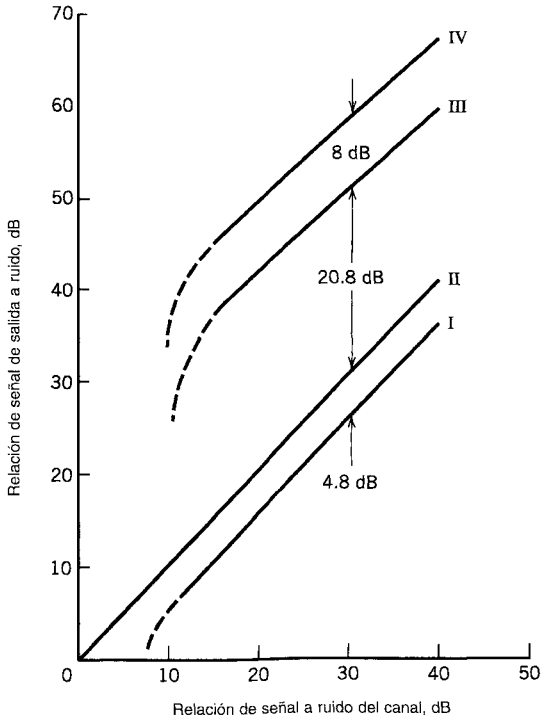


FIGURA 2.55 Comparación del desempeño ante el ruido de diversos sistemas de modulación de onda continua. Curva I: AM completa, $\mu = 1$. Curva II: DSB-SC, SSB. Curva III: FM, $\beta = 2$. Curva IV: FM, $\beta = 5$. (Las curvas III y IV incluyen mejoramiento de preénfasis y deénfasis de 13 dB.)

sistemas de modulación operan exactamente con la misma relación señal a ruido del canal. De ese modo fue posible graficar la relación señal a ruido de salida en función de la relación señal a ruido del canal como en la figura 2.55 para los siguientes esquemas de modulación:

- AM completa con 100% de modulación
- DSB-SC, SSB coherentes
- FM con $\beta = 2$ y $\beta = 5$

La figura 2.55 incluye también los efectos de umbral de la AM y la FM. Al efectuar la comparación, resulta informativo recordar el requerimiento del ancho de banda de transmisión del sistema de modulación respectivo. En consecuencia, utilizamos un *ancho de banda de transmisión normalizado*, definido por

$$B_n = \frac{B_T}{W}$$

donde B_T es el ancho de banda de transmisión de la señal modulada y W es el ancho de banda del mensaje. La tabla 2.4 presenta el valor de B_n para los diferentes esquemas de modulación de onda continua.

A partir de la figura 2.55 y la tabla 2.4 establecemos las siguientes observaciones:

- Entre la familia de sistemas de AM, la modulación de SSB es óptima con respecto al desempeño ante el ruido y la conservación del ancho de banda.
- El uso de la FM mejora el desempeño ante el ruido, aunque a costa de un excesivo ancho de banda de transmisión. Esto supone que el sistema de FM opera sobre el umbral para que se realice el mejoramiento con respecto al ruido.

TABLA 2.4 *Valores de B_n para diversos esquemas de modulación de onda continua*

			FM	
	AM, DSB-SC	*SSB*	$\beta = 2$	$\beta = 5$
B_n	2	1	8	16

Un punto importante para concluir la exposición sobre la modulación de onda continua es el hecho de que sólo la modulación de frecuencia ofrece la capacidad para establecer compromisos entre el ancho de banda de transmisión y un mejor desempeño en relación con el ruido. El compromiso sigue una ley cuadrática, que es lo mejor que podemos lograr con la modulación de onda continua (es decir, las comunicaciones analógicas). En el capítulo 3 describimos la modulación por codificación de pulsos, que es básica para la transmisión de señales analógicas que contienen información por medio de un sistema de comunicación digital, el cual puede en realidad lograr un mucho mejor resultado.

NOTAS Y BIBLIOGRAFÍA

1. Aparentemente los términos *onda continua* y *heterodinado* los utilizó por primera vez Reginald Fessenden a principios del siglo XX.
2. El receptor de Costas recibió ese nombre en honor a su inventor; vea el artículo de Costas [223].
3. Las funciones de Bessel desempeñan un importante papel en el estudio de los sistemas de comunicación tanto analógicos como digitales. Pueden ser de las llamadas de primer o de segundo tipos. En el apéndice 3 se exponen los detalles matemáticos y las propiedades de ambos tipos de funciones de Bessel. En la tabla A6.5 se incluyen funciones de Bessel del primer tipo.
4. La regla de Carson para el ancho de banda de señales de FM tiene ese nombre en honor a su creador; Carson y Fry [213] escribieron uno de los primeros artículos clásicos sobre la teoría de la modulación de frecuencia.
5. El método indirecto para generar una onda de FM de banda ancha fue propuesto por primera vez por Armstrong [188], quien fue el primero en reconocer las propiedades de ruido-robustez de la modulación en frecuencia.
6. El multiplexado estereofónico suele incluir el uso de la modulación de frecuencia para la transmisión de radio. Sin embargo, también es posible transmitirlo utilizando modulación de amplitud como se plantea en el problema 2.14; para mayores detalles, consulte el artículo de Mennie [322].
7. Para una descripción detallada del receptor superheterodino, vea el *Radio Engineering Handbook* editado por Henney ([68], pp. 19-34–19-41).
8. El estudio cualitativo del umbral en la detección de envolvente que se presentó aquí se tomó de Downing ([39], p. 71).
9. En cuanto a una justificación de la suposición crítica relativa al fundamento de la simplificación presentada en la ecuación (2.142), vea Rice [128].
10. Para una exposición detallada del efecto de umbral en los receptores de FM, vea el artículo de Rice [128] y el libro de Schwartz, Bennett y Stein ([133], pp. 129-163).
11. La figura 2.45 se adaptó de Rice [128]. La validez de la curva teórica II en esta figura se ha confirmado experimentalmente; vea Schwartz, Bennett y Stein ([133], p. 153). Para algunos en relación con un trabajo experimental anterior acerca del fenómeno de umbral en FM, vea el artículo de Crosby [225].
12. La idea de utilizar realimentación en torno a un demodulador de FM fue propuesta originalmente por Chaffee [215].

13. El tratamiento del demodulador de FMFB que se presentó en la sección 2.13 se basa en el artículo de Enloe [240]; vea también Roberts ([129], pp. 166-181).
14. Para una explicación detallada de los sistemas Dolby que se mencionaron en la parte final de la sección 2.13, vea Stremler (pp. 732-734).
15. Para un tratamiento completo del análisis no lineal de una malla de fase encadenada, vea Gardner [48], Lindsey [92] y Viterbi [161].

PROBLEMAS

Modulación de amplitud

2.1 Suponga que se cuenta con dispositivos no lineales para los cuales la corriente de salida i_o y el voltaje de entrada v_i se relacionan por medio de

$$i_o = a_1 v_i + a_3 v_i^3$$

donde a_1 y a_3 son constantes. Explique cómo es posible utilizar estos dispositivos para proporcionar: *a*) un modulador de producto y *b*) un modulador de amplitud.

2.2 La figura P2.2 muestra un *diagrama de circuito de un modulador* de ley cuadrática. La señal aplicada al dispositivo no lineal es relativamente débil, de tal modo que puede representarse mediante una ley cuadrática.

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t)$$

donde a_1 y a_2 son constantes, $v_1(t)$ es el voltaje de entrada y $v_2(t)$ es el voltaje de salida. El voltaje de salida se define por medio de

$$v_1(t) = A_c \cos(2\pi f_c t) + m(t)$$

donde $m(t)$ es una señal del mensaje y $A_c \cos(2\pi f_c t)$ es la onda portadora.

a) Evalúe el voltaje de salida $v_2(t)$.

b) Especifique la respuesta en frecuencia que el circuito sintonizado de la figura P2.2 debe satisfacer para generar una señal de AM con f_c como la de la frecuencia de la portadora.

c) ¿Cuál es la sensibilidad de amplitud de esta señal de AM?

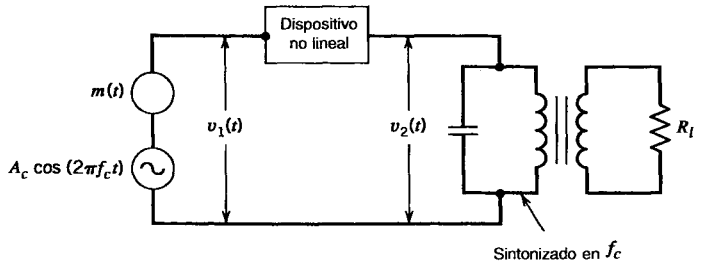


FIGURA P2.2

2.3 La figura P2.3*a* muestra el diagrama de *circuito de un modulador de conmutación*. Suponga que la onda portadora $c(t)$ aplicada al diodo es grande en amplitud, de modo que este último actúa como un interruptor ideal: presenta impedancia cero cuando está polarizado en directo (es decir, $c(t) > 0$).

Por tanto, podemos aproximar la característica de transferencia de la combinación diodo-resistor de carga mediante una característica lineal por secciones definida como (vea la figura P2.3*b*)

$$v_2(t) = \begin{cases} v_1(t), & c(t) > 0 \\ 0, & c(t) < 0 \end{cases}$$

Esto es, el voltaje de carga $v_2(t)$ varía periódicamente entre los valores $v_1(t)$ y cero a una razón igual a la frecuencia de la portadora f_c. Por tanto, podemos escribir

$$v_2(t) \simeq [A_c \cos(2\pi f_c t) + m(t)] g_{T_0}(t)$$

donde $g_{T_0}(t)$ es un tren de pulsos periódico definido por

$$g_{T_0}(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos[2\pi f_c t(2n-1)]$$

a) Encuentre la componente de la onda de AM contenida en el voltaje de salida $v_2(t)$.
b) Especifique las componentes indeseables en $v_2(t)$ que necesitan eliminarse mediante un filtro pasobanda de diseño adecuado.

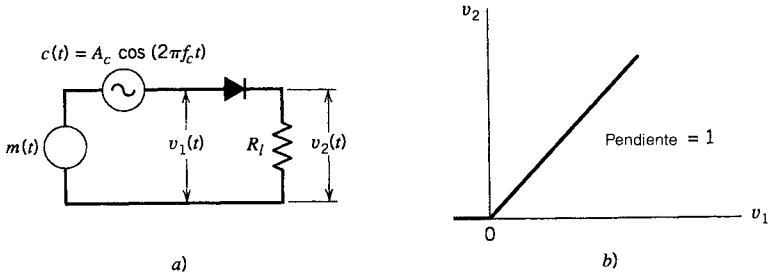


FIGURA P2.3

2.4 Considere la señal de AM

$$s(t) = A_c[1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t)$$

producida por una señal moduladora senoidal de frecuencia f_m. Suponga que el factor de modulación es $\mu = 2$ y que la frecuencia f_c de la portadora es mucho mayor que f_m. La señal $s(t)$ de AM se aplica a un detector de envolvente ideal, produciéndose la salida $v(t)$.

a) Determine la representación en series de Fourier $v(t)$.
b) ¿Cuál es la proporción entre la amplitud de la segunda armónica y la amplitud de la fundamental en $v(t)$?

2.5 La figura P2.5 muestra el diagrama de circuito de un *detector de envolvente*. Éste consiste simplemente en un diodo y un filtro de resistor-capacitor (RC). En el semiciclo positivo de la señal de entrada, el diodo está polarizado en directa y el capacitor C se carga rápidamente hasta el valor pico de la señal de entrada. Cuando la señal de entrada cae por debajo de este valor, el diodo se polariza en inversa y el capacitor C se descarga de manera lenta a través del resistor de carga R_l. El proceso de descarga continúa hasta el siguiente semiciclo positivo. De ahí en adelante, continúa la rutina de carga-descarga.

a) Especifique la condición que debe cumplir el capacitor C para cargarlo rápidamente y seguir de ese modo el voltaje de entrada ascendente hasta el pico positivo cuando está conduciendo el diodo.

b) Especifique la condición que el resistor de carga R_l debe satisfacer de manera que el capacitor C se descargue en forma lenta entre los picos positivos de la onda portadora, pero no tanto

como para que el voltaje del capacitor no se descargue a la velocidad máxima de cambio de la onda moduladora.

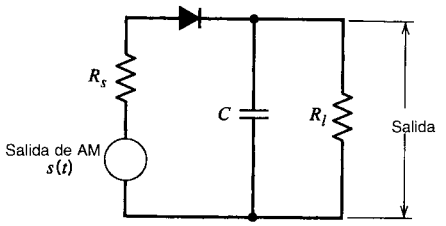


FIGURA P2.5

2.6 Considere un *detector de ley cuadrática*, utilizando un dispositivo no lineal cuya característica de transferencia está definida por

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t)$$

donde a_1 y a_2 son constantes, $v_1(t)$ es la entrada y $v_2(t)$ es la salida. La entrada consiste en una onda de AM

$$v_1(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$

a) Evalúe la salida $v_2(t)$.

b) Encuentre las condiciones para la cuales la señal de mensaje $m(t)$ pueda recuperarse de $v_2(t)$.

2.7 La señal de AM

$$s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$

se aplica al sistema que se presenta en la figura P2.7. Suponiendo que $|k_a m(t)| < 1$ para toda t y que la señal de mensaje $m(t)$ está limitada al intervalo $-W \leq f \leq W$ y que la frecuencia de la portadora $f_c > 2W$, demuestre que $m(t)$ puede obtenerse de la salida del extractor de raíz cuadrada $v_3(t)$.

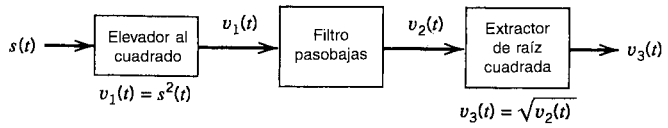


FIGURA P2.7

2.8 Considere una señal de mensaje $m(t)$ con el espectro que se muestra en la figura P2.8. El ancho de banda del mensaje corresponde a $W = 1$ kHz. Esta señal se aplica a un modulador de producto, junto con una onda portadora $A_c \cos(2\pi f_c t)$, que produce la señal modulada de DSB-SC $s(t)$. La señal modulada se aplica después a un detector coherente. Suponiendo sincronismo perfecto entre las ondas portadoras en el modulador y en el detector, determine el espectro de la salida del detector cuando: *a)* la frecuencia de la portadora $f_c = 1.25$ kHz, y *b)* la frecuencia de la portadora $f_c = 0.75$ kHz. ¿Cuál es la frecuencia de portadora más baja para la cual cada componente de la señal modulada $s(t)$ está determinada en forma única por $m(t)$?

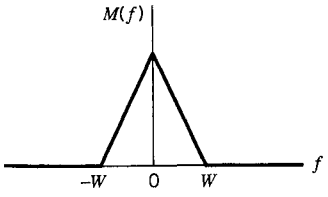


FIGURA P2.8

2.9 La figura P2.9 muestra el diagrama de circuito de un *modulador balanceado*. La entrada aplicada a la parte superior del modulador de AM es $m(t)$, en tanto que la que se aplica al modulador de AM inferior es $-m(t)$; estos dos moduladores tienen la misma sensibilidad de amplitud. Demuestre que la salida $s(t)$ del modulador balanceado consiste en una señal modulada de DSB-SC.

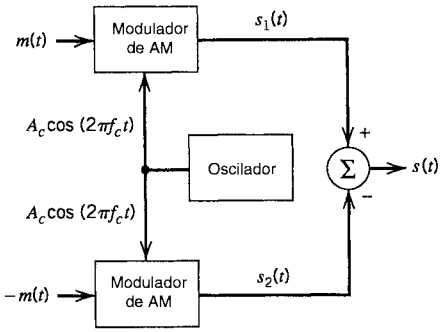


FIGURA P2.9

2.10 Una señal modulada de DSB-SC se demodula aplicándola a un detector coherente.

a) Evalúe el efecto de un error de frecuencia Δf en la frecuencia de la portadora local del detector, medida con respecto a la frecuencia de la portadora de la señal de DSB-SC entrante.

b) Para el caso de una onda moduladora senoidal, demuestre que debido a este error de frecuencia, la señal demodulada presenta *pulsaciones* a la frecuencia de error. Ilustre su respuesta con un dibujo de esta señal demodulada.

2.11 Considere la señal de DSB-SC

$$s(t) = A_c \cos(2\pi f_c t) m(t)$$

donde $A_c \cos(2\pi f_c t)$ es la onda portadora y $m(t)$ es la señal de mensaje. Esta señal modulada se aplica a un dispositivo de ley cuadrática caracterizado por

$$y(t) = s^2(t)$$

La salida $y(t)$ se aplica luego a un filtro de banda angosta con una respuesta en magnitud pasobanda de uno, frecuencia de banda media $2f_c$, y ancho de banda Δf. Suponga que Δf es lo bastante pequeña para tratar el espectro de $y(t)$ como esencialmente constante dentro de la banda de paso del filtro.

a) Determine el espectro de la salida $y(t)$ del dispositivo de ley cuadrática.

b) Muestre que la salida $v(t)$ del filtro es aproximadamente senoidal y que está dada por

$$v(t) \simeq \frac{A_c^2}{2} E \Delta f \cos(4\pi f_c t)$$

donde E es la energía de la señal del mensaje $m(t)$.

2.12 Considere el sistema de multiplexado de portadora en cuadratura de la figura 2.10. La señal de multiplexado $s(t)$ que se produce a la salida del transmisor en la figura 2.10*a* se aplica a un canal de comunicación de respuesta en frecuencia $H(f)$. La salida de este canal se aplica, a su vez, a la entrada del receptor de la figura 2.10*b*. Demuestre que la condición

$$H(f_c + f) = H^*(f_c - f), \quad 0 \le f \le W$$

es necesaria para recuperar las señales del mensaje $m_1(t)$ y $m_2(t)$ a la salida del receptor; f_c es la frecuencia de la portadora y W es el ancho de banda del mensaje. *Sugerencia:* evalúe el espectro de las dos salidas del receptor.

2.13 Suponga que en el receptor del sistema de multiplexado de portadora en cuadratura de la figura 2.10*b* la portadora local disponible para la demodulación tiene un error de fase ϕ con respecto a la fuente de la portadora que se usa en el transmisor. Suponga un canal de comunicaciones sin distorsión entre el transmisor y el receptor, muestre que este error de fase ocasionará que se produzca *diafonía* entre las dos señales demoduladas a las salidas del receptor. Por diafonía damos a entender que una parte de una señal del mensaje aparece en la salida del receptor que pertenece a la otra señal de mensaje, y viceversa.

2.14 Una versión particular de un *estereofónico de* AM utiliza multiplexado en cuadratura. Específicamente, la portadora $A_c \cos(2\pi f_c t)$ se usa para modular la señal suma

$$m_1(t) = V_0 + m_\ell(t) + m_r(t)$$

donde V_0 es una compensación de cd que se incluye con el fin de transmitir la componente de la portadora, $m_\ell(t)$ es la señal de audio del lado izquierdo y $m_r(t)$ es la señal del audio de lado derecho. La portadora en cuadratura $A_c \operatorname{sen}(2\pi f_c t)$ se utiliza para modular la señal de diferencia

$$m_2(t) = m_\ell(t) - m_r(t)$$

a) Demuestre que es posible utilizar un detector de envolvente para recuperar la suma $m_r(t) + m_\ell(t)$ de la señal multiplexada en cuadratura. ¿Cómo minimizaría usted la distorsión de señal que produce el detector de envolvente?

b) Demuestre que un detector coherente puede recuperar la diferencia $m_\ell(t) - m_r(t)$.

c) ¿Cómo se obtienen finalmente las $m_\ell(t)$ y $m_r(t)$ deseadas?

2.15 Utilizando la señal de mensaje

$$m(t) = \frac{1}{1+t^2}$$

determine y bosqueje las ondas moduladas en los siguientes métodos de modulación:

a) Modulación de amplitud con 50% de modulación.

b) Modulación de doble banda lateral-portadora suprimida.

c) Modulación de una sola banda lateral con únicamente la banda lateral superior transmitida.

d) Modulación de banda lateral única con únicamente la banda lateral inferior transmitida.

2.16 La *transformada de Hilbert* de una señal transformable de Fourier $m(t)$, denotada por $\hat{m}(t)$, se define por medio de (vea el apéndice 2)

$$\hat{m}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{m(\tau)}{t-\tau} d\tau$$

En el dominio de la frecuencia, tenemos

$$\hat{M}(f) = -j \operatorname{sgn}(f) M(f)$$

donde $m(t) \rightleftharpoons M(f)$, $\hat{m}(t) \rightleftharpoons \hat{M}(f)$, y $\operatorname{sgn}(f)$ es la función signo.

Empleando la definición de la transformada de Hilbert, demuestre que una señal modulada de banda lateral única a partir de la señal de mensaje $m(t)$ y la portadora $\cos(2\pi f_c t)$ de amplitud unitaria está dada por (vea la tabla 2.1)

$$s(t) = \frac{1}{2}m(t)\cos(2\pi f_c t) \pm \frac{1}{2}\hat{m}(t)\,\text{sen}(2\pi f_c t)$$

donde el signo menos corresponde a la transmisión de la banda lateral superior y el signo más a la transmisión de la banda lateral inferior.

2.17 El oscilador local que se usa para la demodulación de una señal de SSB $s(t)$ tiene un error de frecuencia Δf medido con respecto a la frecuencia de la portadora f_c que se utiliza para generar $s(t)$. De otro modo, hay un sincronismo perfecto entre este oscilador en el receptor y el oscilador que suministra la onda portadora en el transmisor. Evalúe la señal demodulada en las siguientes dos situaciones:

a) La señal de SSB $s(t)$ está compuesta sólo por la banda lateral superior.

b) La señal de SSB $s(t)$ consiste sólo en la banda lateral inferior.

2.18 La figura P2.18 muestra el diagrama de bloques del *método de Weaver* para generar ondas moduladas de SSB. La señal de mensaje (moduladora) $m(t)$ se limita a la banda $f_a \leq |f| \leq f_b$. La portadora auxiliar aplicada al primer par de moduladores de producto tiene una frecuencia f_0 que yace en el centro de esta banda, según indica

$$f_0 = \frac{f_a + f_b}{2}$$

Los filtros pasobajas en los canales en fase y en cuadratura son idénticos, cada uno con una frecuencia de corte igual a $(f_b - f_a)/2$. La portadora aplicada al segundo par de moduladores de producto tiene una frecuencia f_c que es mayor a $(f_b - f_a)/2$. Dibuje los espectros en los diversos puntos en el modulador de la figura P2.18 y, por tanto, demuestre que:

a) Para la banda lateral inferior, las contribuciones de los canales en fase y en cuadratura son de polaridad opuesta, y sumándolos a la salida del modulador, se suprime la banda lateral inferior.

b) Para la banda lateral superior, las contribuciones de los canales en fase y en cuadratura son de la misma polaridad, y sumándolos, se transmite la banda lateral superior.

c) ¿Cómo modificaría el modulador de la figura P2.18 de manera que sólo se transmita la banda lateral inferior?

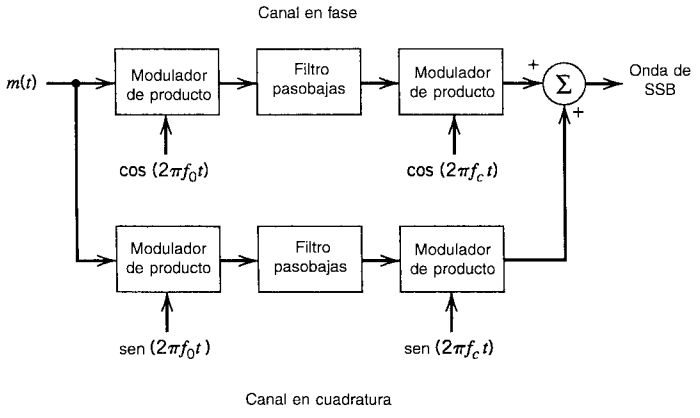


FIGURA P2.18

2.19 El espectro de una señal de voz $m(t)$ es cero fuera del intervalo $f_a \leq |f| \leq f_b$. Para asegurar la privacía de comunicación, esta señal se aplica a un *criptógrafo* que consiste en los siguientes componentes en cascada: modulador de producto, filtro pasoaltas, segundo modulador de producto y filtro pasobajas. La onda portadora aplicada al primer modulador de producto tiene una frecuencia igual a f_c, en tanto que la del segundo modulador de producto corresponde a $f_b + f_c$; ambos tienen amplitud unitaria. Los filtros pasoaltas y pasobajas tienen la misma frecuencia de corte en f_c. Suponga que $f_c > f_b$.

a) Obtenga una expresión para la salida de $s(t)$ del criptógrafo y dibuje su espectro.

b) Demuestre que la señal de voz original $m(t)$ puede recuperarse a partir de $s(t)$ utilizando un *descriptógrafo* que es idéntico a la unidad que acaba de describirse.

2.20 En este problema obtenemos la ecuación (2.16) que define la respuesta en frecuencia $H_Q(f)$ del filtro que opera en la señal de mensaje $m(t)$ para producir la $m'(t)$ correspondiente a la modulación de VSB. La señal $m'(t)$, salvo por un factor de escalamiento, constituye la componente en cuadratura de $s(t)$. Para llevar a cabo la deducción, aplicamos $s(t)$ al detector coherente de la figura P2.20 de modo que se recupere la versión escalada de la señal de mensaje original $m(t)$.

a) Empezando con el diagrama de bloques de la figura 2.12 para la generación de una onda modulada de VSB determine la transformada de Fourier $V(f)$ de la salida $v(t)$ del modulador de producto en la figura P2.20 en términos de la transformada de Fourier de la señal de mensaje $m(t)$ y de la respuesta en frecuencia $H(f)$ del filtro pasobanda de la figura 2.12.

b) Por tanto, al evaluar la transformada de Fourier de la salida del filtro pasobajas de la figura P2.20, determine las condiciones que debe cumplir $H_Q(f)$ en términos de $H(f)$ para asegurar la recuperación perfecta de la señal de mensaje original $m(t)$, excepto por un factor de escalamiento.

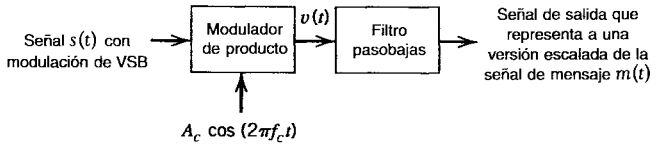


FIGURA P2.20

2.21 La señal moduladora de un solo tono $m(t) = A_m \cos(2\pi f_m t)$ se usa para generar la señal VSB

$$s(t) = \frac{1}{2} a A_m A_c \cos[2\pi(f_c + f_m)t] + \frac{1}{2} A_m A_c (1-a) \cos[2\pi(f_c - f_m)t]$$

donde a es una constante, menor que la unidad, que representa la atenuación de la frecuencia lateral superior.

a) Determine la componente en cuadratura de la señal de VSB $s(t)$.

b) La señal de VSB, más la portadora $A_c \cos(2\pi f_c t)$ se hace pasar por un detector de envolvente. Determine la distorsión que produce la componente en cuadratura.

c) ¿Cuál es el valor de la constante a con la cual esta distorsión alcanza su peor condición posible?

2.22 En este problema estudiamos la idea de mezcla en un receptor superheterodino. De manera específica, considere el diagrama de bloques del *mezclador* que se presenta en la figura P2.22 y que consiste en un modulador de producto con un oscilador local de *frecuencia variable* f_l seguido por un filtro pasobanda. La señal de entrada es una onda de AM de ancho de banda igual a 10 kHz y una frecuencia de portadora que puede estar dentro del intervalo de 0.535 a 1.605 MHz; estos parámetros son característicos de la transmisión de radio de AM. Se requiere trasladar esta señal

a una banda de frecuencia centrada en una *frecuencia intermedia* (FI) fija de 0.455 MHz. Determine el intervalo de sintonía que debe proporcionarse en el oscilador local para lograr este requerimiento.

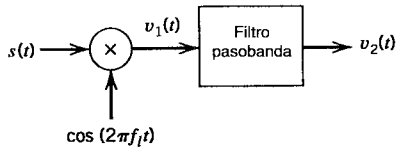


FIGURA P2.22

2.23 La figura P2.23 muestra el diagrama de bloques de un *analizador de espectros heterodino*. Éste consiste en un oscilador de frecuencia variable, multiplicador, filtro pasobanda y medidor de valor medio cuadrático (rms). El oscilador tiene una amplitud A y opera sobre el intervalo de f_0 a $f_0 + W$, donde f_0 es la frecuencia de banda media del filtro y W es el ancho de banda de la señal. Suponga que $f_0 = 2W$, el ancho de banda del filtro Δf es pequeño en comparación con f_0 y que la respuesta en magnitud de la banda de paso del filtro es igual a uno. Determine el valor de la salida del medidor de rms para una señal de entrada pasobajas $g(t)$.

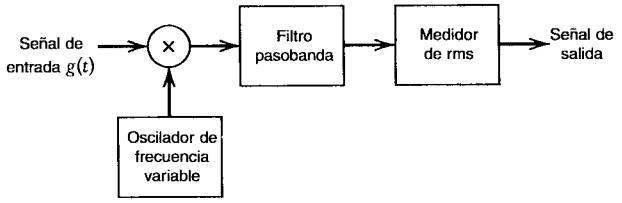


FIGURA P2.23

Modulación angular

2.24 Dibuje las ondas de PM y FM producidas por la onda de diente de sierra que se muestra en la figura P2.24.

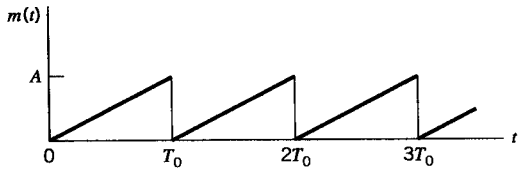


FIGURA P2.24

2.25 En un *radar modulado por frecuencia*, la frecuencia instantánea de la portadora transmitida se varía como en la figura P2.25, la cual se obtiene al utilizar una señal moduladora triangular. La frecuencia instantánea de la señal de eco recibida se muestra mediante la línea punteada en la figura P2.25, donde τ es el tiempo de retardo del viaje redondo. Las señales del eco transmitido y recibido se aplican a un mezclador, y se retiene la componente de diferencia de frecuencia. Suponiendo que $f_0\tau \ll 1$, determine el número de ciclos de pulsación en la salida del mezclador, promediados

sobre un segundo, en términos de la desviación pico Δf de la frecuencia de portadora, el retraso τ y la frecuencia de repetición f_0 de la señal transmitida.

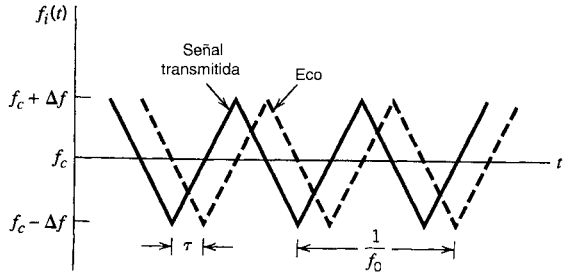


FIGURA P2.25

2.26 La frecuencia instantánea de una onda seno es igual a $f_c - \Delta f$ para $|t| \leq T/2$, y f_c para $|t| > T/2$. Determine el espectro de esta onda modulada en frecuencia. *Sugerencia*: divida el intervalo de interés en tres regiones: $-\infty < t < -T/2$, $-T/2 \leq t \leq T/2$ y $T/2 < t < \infty$.

2.27 La modulación de banda lateral única puede considerarse como una forma híbrida de modulación de amplitud y de modulación de frecuencia. Evalúe la envolvente en frecuencia instantánea de una onda de SSB para los siguientes dos casos:

a) Cuando sólo se transmite la banda lateral superior.

b) Cuando sólo se transmite la banda lateral inferior.

2.28 Considere una señal de FM de banda angosta definida aproximadamente por

$$s(t) \simeq A_c \cos(2\pi f_c t) - \beta A_c \operatorname{sen}(2\pi f_c t) \operatorname{sen}(2\pi f_m t)$$

a) Determine la envolvente de esta señal modulada. ¿Cuál es la proporción entre los valores máximo y mínimo de esta envolvente? Grafique esta proporción en función de β, suponiendo que β está restringida al intervalo $0 \leq \beta \leq 0.3$.

b) Determine la potencia promedio de una señal de FM de banda angosta, expresada como un porcentaje de la potencia promedio de la onda portadora no modulada. Grafique este resultado en función de β, suponiendo que β está restringida al intervalo $0 \leq \beta \leq 0.3$.

c) Expandiendo el ángulo $\theta_i(t)$ de la señal de FM de banda angosta $s(t)$ en la forma de una serie de potencias y restringiendo el índice de modulación β a un valor máximo de 0.3 radianes, demuestre que

$$\theta_i(t) \simeq 2\pi f_c t + \beta \operatorname{sen}(2\pi f_m t) - \frac{\beta^3}{3} \operatorname{sen}^3(2\pi f_m t)$$

¿Cuál es la proporción de potencia de la tercera armónica con la componente fundamental para $\beta = 0.3$?

2.29 La onda moduladora senoidal

$$m(t) = A_m \cos(2\pi f_m t)$$

se aplica a un modulador de fase con sensibilidad de fase k_p. La onda portadora no modulada tiene frecuencia f_c y amplitud A_c.

a) Determine el espectro de la señal modulada en fase que resulta, suponiendo que la desviación de fase máxima $\beta_p = k_p A_m$ no excede de 0.3 radianes.

b) Construya un diagrama fasorial para esta señal modulada y compárela con la de la señal de FM de banda angosta correspondiente.

2.30 Suponga que la señal modulada en fase del problema 2.29 tiene un valor arbitrario para la máxima desviación de fase β_p. Esta señal modulada se aplica a un filtro pasobanda ideal con frecuencia de banda media f_c y a una banda de paso que se extiende de $f_c - 1.5f_m$ a $f_c + 1.5f_m$. Determine la

envolvente, fase y frecuencia instantánea de la señal modulada a la salida del filtro como una función del tiempo.

2.31 Una onda portadora se modula en frecuencia utilizando una señal senoidal de frecuencia f_m y amplitud A_m.

a) Determine el valor del índice de modulación β para el cual la componente de la portadora de la señal de FM se reduce a cero. En este cálculo usted puede utilizar los valores de $J_0(\beta)$ que se dan en la tabla A6.5.

b) En cierto experimento conducido con $f_m = 1$ kHz y A_m creciente (empezando desde 0 volt), se encuentra que la componente de la portadora de la señal de FM se reduce a cero en el primer momento en el que $A_m = 2$ volts. ¿Cuál es la sensibilidad de frecuencia del modulador? ¿Cuál es el valor de A_m para el cual las componentes de la portadora se reducen a cero en el segundo momento?

2.32 Una señal de FM con índice de modulación $\beta = 1$ se transmite por un filtro pasobanda ideal con frecuencia de banda media f_c y ancho de banda $5f_m$, donde f_c es la frecuencia de la portadora y f_m es la frecuencia de la onda moduladora senoidal. Determine el espectro de magnitud de la salida del filtro.

2.33 Una onda portadora de 100 MHz de frecuencia se modula en frecuencia por medio de una onda senoidal de 20 volts de amplitud y 100 kHz de frecuencia. La sensibilidad de frecuencia del modulador es de 25 kHz por volt.

a) Determine el ancho de banda aproximado de la señal de FM, utilizando la regla de Carson.

b) Determine el ancho de banda al transmitir sólo aquellas frecuencias laterales cuya amplitud supera 1% de la amplitud de la portadora sin modulación. Utilice la curva universal de la figura 2.26 para este cálculo.

c) Repita sus cálculos, suponiendo que se duplica la amplitud de la señal moduladora.

d) Repita sus cálculos, suponiendo que se duplica la frecuencia de modulación.

2.34 Considere una señal de PM de banda ancha producida por una onda moduladora senoidal $A_m \cos(2\pi f_m t)$, utilizando un modulador con una sensibilidad de fase igual a k_p radianes por volt.

a) Demuestre que si la desviación de fase máxima de la señal de PM es grande en comparación con un radián, el ancho de banda de la señal de PM varía linealmente con la frecuencia de modulación f_m.

b) Compare esta característica de la señal de PM de banda ancha con la de una señal de FM de banda ancha.

2.35 La figura P2.35 muestra el diagrama de bloques de un *analizador de espectros* en tiempo real que funciona con base en el principio de modulación de frecuencia. La señal dada $g(t)$ y la señal modulada en frecuencia $s(t)$ se aplican a un multiplicador y la salida $g(t)s(t)$ se alimenta a un filtro de respuesta al impulso $h(t)$. $s(t)$ y $h(t)$ son *señales de FM lineales* cuyas frecuencias instantáneas varían linealmente con el tiempo a velocidades opuestas, de acuerdo con

$$s(t) = \cos(2\pi f_c t - \pi k t^2)$$
$$h(t) = \cos(2\pi f_c t - \pi k t^2)$$

donde k es una constante. Demuestre que la envolvente de la salida del filtro es proporcional al espectro de magnitud de la señal de entrada $g(t)$ con kt desempeñando el papel de la frecuencia f. *Sugerencia:* use la anotación compleja que se describe en el apéndice 2 para el análisis de las señales pasobanda y los filtros pasobanda.

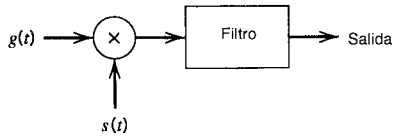


FIGURA P2.35

2.36 Una señal de FM con una desviación de frecuencia de 10 kHz a una frecuencia de modulación de 5 kHz se aplica a dos multiplicadores de frecuencia conectados en cascada. El primer multiplicador duplica la frecuencia y el segundo la triplica. Determine la desviación de frecuencia y el índice de modulación de la señal de FM que se obtiene en la segunda salida del multiplicador. ¿Cuál es la separación de frecuencia de las frecuencias laterales adyacentes de esta señal de FM?

2.37 La figura P2.37 muestra el diagrama de bloques de un modulador de frecuencia de banda ancha que utiliza el método indirecto. Este modulador se utiliza para transmitir señales de audio que contienen frecuencias en el intervalo de 100 Hz a 15 kHz. El modulador de fase de banda angosta se alimenta con una portadora de frecuencia $f_1 = 0.1$ MHz mediante un oscilador controlado por cristal. Un segundo oscilador controlado por cristal suministra una onda senoidal de 9.5 MHz de frecuencia al mezclador. Las especificaciones del sistema son las siguientes:

Frecuencia de la portadora a la salida del transmisor, $f_c = 100$ MHz

Desviación de frecuencia mínima, $\Delta f = 75$ kHz

Índice de modulación máxima en el modulador de fase = 0.3 radianes

a) Calcule las proporciones de multiplicación de frecuencia n_1 y n_2 (precediendo y siguiendo al mezclador) que satisfarán estas especificaciones.

b) Especifique los valores de la frecuencia de la portadora y la desviación de frecuencia en diversos puntos en el modulador de la figura P2.37.

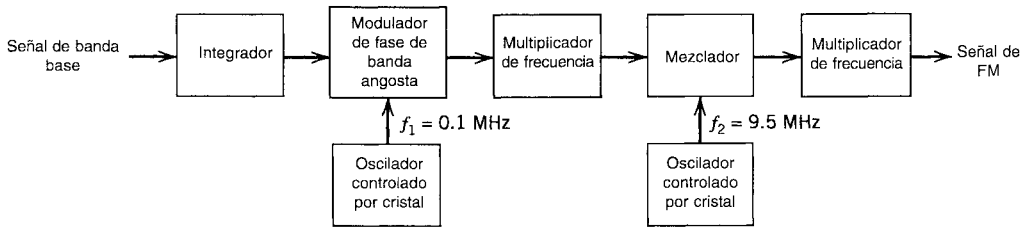


FIGURA P2.37

2.38 La figura P2.38 muestra la red de determinación de frecuencia de un oscilador controlado por voltaje. La modulación de frecuencia se produce aplicando la señal de modulación $A_m \operatorname{sen}(2\pi f_m t)$ más una polarización V_b a un par de diodos varactores conectados a través de la combinación paralela de un inductor de 200 μH y un capacitor de 100 pF. El capacitor de cada diodo varactor se relaciona con el voltaje V (en volt) aplicado en sus electrodos por medio de

$$C = 100V^{-1/2} \text{ pF}$$

La frecuencia no modulada de oscilación es 1 MHz. La salida del oscilador controlado por voltaje se aplica a un multiplicador de frecuencia para producir una señal de FM con una frecuencia de portadora de 64 MHz y un índice de modulación de 5. Determine *a)* la magnitud del voltaje de polarización V_b y *b)* la amplitud A_m de la onda moduladora, dado que $f_m = 10$ kHz.

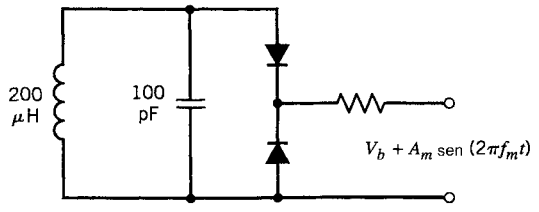


FIGURA P2.38

2.39 La señal de FM

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\, d\tau\right]$$

se aplica al sistema que se muestra en la figura P2.39 consistente en un filtro RC pasoaltas y un detector de envolvente. Suponga que *a*) la resistencia R es pequeña en comparación con la reactancia del capacitor C para todos los componentes de frecuencia significativas de $s(t)$, y *b*) el detector de envolvente no carga el filtro. Determine la señal resultante a la salida del detector de envolvente, suponiendo que $k_f |m(t)| < f_c$ para toda t.

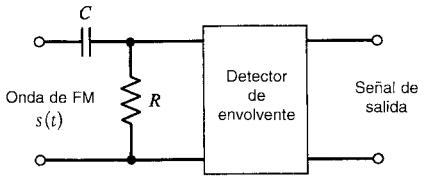


FIGURA P2.39

2.40 Considere el esquema de demodulación de frecuencia que se muestra en la figura P2.40 en el cual la señal de FM entrante $s(t)$ se hace pasar por una línea de retardo que produce un corrimiento de fase de $\pi/2$ radianes a la frecuencia de portadora f_c. La salida de la línea de retardo se resta de la señal de FM entrante, y la señal compuesta que resulta se detecta entonces en la envolvente. Este demodulador encuentra aplicación en la demodulación de señales de FM de microondas. Suponiendo que

$$s(t) = A_c \cos[2\pi f_c t + \beta \operatorname{sen}(2\pi f_m t)]$$

analice la operación de este demodulador cuando el índice de modulación β es menor que la unidad y el retardo T producido por la línea de retraso es suficientemente pequeña para justificar las aproximaciones

$$\cos(2\pi f_m T) \simeq 1$$

y

$$\operatorname{sen}(2\pi f_m T) \simeq 2\pi f_m T$$

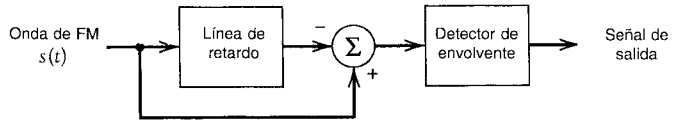


FIGURA P2.40

2.41 La figura P2.41 muestra el diagrama de bloques de *un detector de cruce por cero* para demodular una señal de FM. Está compuesto por un limitador, un generador de pulsos para producir un pulso corto en cada cruce por cero de la entrada, y un filtro pasobajas para extraer la onda moduladora.

a) Demuestre que la frecuencia instantánea de la señal de FM de entrada es proporcional al número de cruces por cero en el intervalo de tiempo $t - (T_1/2)$ a $t + (T_1/2)$, dividido entre T_1. Suponga que la señal moduladora es esencialmente constante durante este intervalo de tiempo.

b) Ilustre la operación de este demodulador, utilizando la onda diente de cierra de la figura P2.24 como la onda moduladora.

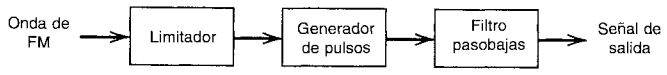


FIGURA P2.41

2.42 Suponga que la señal recibida en un sistema de FM contiene alguna modulación de amplitud residual de amplitud positiva $a(t)$, según se indica mediante

$$s(t) = a(t)\cos[2\pi f_c t + \phi(t)]$$

donde f_c es la frecuencia de la portadora. La fase $\phi(t)$ se relaciona con la señal moduladora $m(t)$ por

$$\phi(t) = 2\pi k_f \int_0^t m(\tau)\, d\tau$$

donde k_f es una constante. Suponga que la señal $s(t)$ se restringe a una banda de frecuencia de ancho B_T, centrada en f_c, donde B_T es el ancho de banda de transmisión de la señal de FM ante la ausencia de modulación de amplitud, y que la modulación de amplitud está variando lentamente en comparación con $\phi(t)$. Demuestre que la salida de un discriminador de frecuencia ideal que produce $s(t)$ es proporcional a $a(t)m(t)$. *Sugerencia:* use la notación completa descrita en el apéndice 2 para representar la onda modulada $s(t)$.

2.43 a) Considere que la onda modulada $s(t)$ del problema 2.42 se aplica a un *limitador rígido*, cuya salida $z(t)$ está definida por

$$z(t) = \text{sgn}[s(t)] = \begin{cases} +1, & s(t) > 0 \\ -1, & s(t) < 0 \end{cases}$$

Demuestre que la salida del limitador puede expresarse en la forma de una serie de Fourier del modo siguiente:

$$z(t) = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\cos[2\pi f_c t(2n+1) + (2n+1)\phi(t)]$$

b) Suponga que la salida del limitador se aplica a un filtro pasobanda con una respuesta en magnitud pasobanda de uno y ancho de banda B_T centrada en torno a la frecuencia de la portadora f_c, donde B_T es el ancho de banda de transmisión de la señal de FM en la ausencia de modulación de amplitud. Suponiendo que f_c es mucho mayor que B_T, demuestre que la salida del filtro resultante es igual a

$$y(t) = \frac{4}{\pi}\cos[2\pi f_c t + \phi(t)]$$

Al comparar esta salida con la señal modulada original $s(t)$ definida en el problema 2.42, comente acerca de la utilidad práctica del resultado.

2.44 a) Considere una señal de FM de frecuencia de portadora f_c, que produce una señal moduladora $m(t)$. Suponga que f_c es suficientemente grande para justificar el trato de esta señal de FM como una señal de banda angosta. Determine una expresión aproximada para esta transformada de Hilbert.

b) Para el caso especial de una onda moduladora senoidal $m(t) = A_m \cos(2\pi f_m t)$, encuentre la expresión exacta para la transformada de Hilbert de la señal de FM resultante. En este caso, ¿cuál es el error de la aproximación que se utilizó en la parte *a*?

2.45 La *versión de banda lateral única de modulación angular* se define por medio de

$$s(t) = \exp\left[-\hat{\phi}(t)\right] \cos\left[2\pi f_c t + \phi(t)\right]$$

donde $\hat{\phi}(t)$ es la transformada de Hilbert de la función de fase $\phi(t)$, y f_c es la frecuencia de la portadora.

a) Demuestre que el espectro de la señal modulada $s(t)$ no contiene componente de frecuencia en el intervalo $-f_c < f < f_c$, y que es de extensión infinita.

b) Dado que la función de fase

$$\phi(t) = \beta \operatorname{sen}(2\pi f_m t)$$

donde β es el índice de modulación y f_m es la frecuencia de modulación, obtenga las expresiones correspondientes para la onda modulada $s(t)$.

Nota: en los problemas 2.44 y 2.45 necesita referirse al apéndice 2 para un tratamiento de la transformada de Hilbert.

Ruido en sistemas de modulación de onda continua

2.46 Una señal modulada de DSB-SC se transmite sobre un canal ruidoso, siendo la densidad espectral de potencia del ruido como se indica en la figura P2.46. El ancho de banda del mensaje corresponde a 4 kHz y la frecuencia de la portadora es 200 kHz. Suponiendo que la potencia promedio de la onda modulada es igual a 10 watts, determine la relación señal a ruido de salida del receptor.

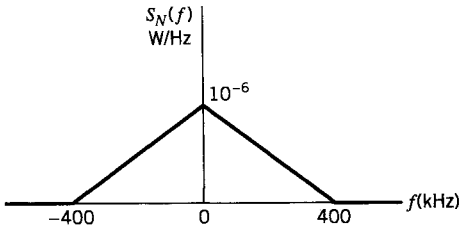


FIGURA P2.46

2.47 Evalúe las funciones de autocorrelación y las funciones de correlación cruzada de las componentes en fase y en cuadratura del ruido de banda angosta a la entrada del detector coherente de *a*) el sistema de DSB-SC, *b*) un sistema de SSB utilizando la banda lateral inferior y *c*) un sistema SSB utilizando la banda lateral superior.

2.48 En un receptor que utiliza detección coherente, la onda senoidal que genera el oscilador local sufre un error de fase $\theta(t)$ con respecto a la onda portadora $\cos(2\pi f_c t)$. Suponiendo que $\theta(t)$ es una función de muestreo de un proceso gaussiano de media cero y varianza σ_Θ^2 y que la mayor parte del tiempo el valor máximo de $\theta(t)$ es pequeño en comparación con la unidad, encuentre el error cuadrático medio de la salida del receptor para modulación de DSB-CS. El error cuadrático medio se define como el valor esperado de la diferencia al cuadrado entre la salida del receptor y la componente de la señal de mensaje de la salida del receptor.

2.49 Siguiendo un procedimiento similar al que se describió en la sección 2.11 para el receptor de DSB-CS, extienda este análisis de ruido a un receptor de SSB. De manera específica, evalúe lo siguiente:

a) La relación señal de salida a ruido de salida.

b) La relación señal a ruido del canal.

Por consiguiente, demuestre que la figura de mérito para un receptor de SSB es exactamente la misma que para el receptor de DSB-SC. Advierta que a diferencia del receptor DSB-SC, la frecuencia de banda media de la función de densidad espectral del ruido filtrado de banda angosta en el extremo frontal del receptor de SSB está descompensada con respecto a la frecuencia de portadora f_c en una cantidad igual a $W/2$, donde W es el ancho de banda del mensaje.

2.50 Considere que una señal de mensaje $m(t)$ se transmite utilizando modulación de banda lateral única. La densidad espectral de potencia de $m(t)$ es

$$S_M(f) = \begin{cases} a\dfrac{|f|}{W}, & |f| \le W \\ 0, & \text{en otro caso} \end{cases}$$

donde a y W son constantes. Se agregan ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$ a la onda modulada de SSB a la entrada del receptor. Encuentre una expresión para la relación señal a ruido de salida del receptor.

2.51 Considere la salida de un detector de envolvente definida por la ecuación (2.92), que se repite aquí por conveniencia

$$y(t) = \left\{\left[A_c + A_c k_a m(t) + n_I(t)\right]^2 + n_Q^2(t)\right\}^{1/2}$$

a) Suponga que la probabilidad del evento

$$\left|n_Q(t)\right| > \varepsilon A_c \left|1 + k_a m(t)\right|$$

es igual o menor que δ_1 donde $\varepsilon << 1$. ¿Cuál es la probabilidad de que sea ignorable el efecto de la componente en cuadratura $n_Q(t)$?

b) Suponga que k_a se ajusta en relación con la señal de mensaje $m(t)$ de manera tal que la probabilidad del evento

$$A_c\left[1 + k_a m(t)\right] + n_I(t) < 0$$

es igual a δ_2. ¿Cuál es la probabilidad de que la aproximación

$$y(t) \simeq A_c\left[1 + k_a m(t)\right] + n_I(t)$$

sea válida?

c) Comente acerca de la importancia del resultado en la parte *b* para el caso en el que δ_1 y δ_2 son ambas pequeñas en comparación con la unidad.

2.52 Una portadora no modulada en amplitud A_c y frecuencia f_c y ruido blanco limitado en banda, se suman y luego se pasan por un detector de envolvente ideal. Suponga que la densidad espectral de ruido será de altura $N_0/2$ y ancho de banda $2W$, centrada en torno a la frecuencia de la portadora f_c. Determine la relación señal a ruido de salida para el caso en el que la relación de portadora a ruido es elevada.

2.53 Deje que R denote la variable aleatoria que se obtiene al observar la salida de un detector de envolvente a cierto tiempo fijo. Intuitivamente, se espera que el detector de envolvente opere bien dentro de la región de umbral si la probabilidad de que la variable aleatoria R supere la amplitud de la portadora A_c es de 0.5. Por otra parte, si esta misma probabilidad es únicamente de 0.01, se espera que el detector de envolvente no resulte alterado por la pérdida de mensaje y el efecto de umbral.

a) Suponiendo que el ruido de banda angosta en la entrada del detector es blanco gaussiano de media cero con densidad espectral $N_0/2$ y el ancho de banda del mensaje es W, demuestre que la probabilidad del evento $R \ge A_c$ es

$$P(R \ge A_c) = \exp(-\rho)$$

donde ρ es la relación portadora a ruido:

$$\rho = \frac{A_c^2}{4WN_0}$$

b) Utilizando la fórmula para esta probabilidad, calcule la relación portadora a ruido cuando: 1) se espera que el detector de envolvente se encuentre por completo en la región de umbral y 2) se tiene la certeza de que operará en forma satisfactoria.

2.54 Considere el sistema de modulación de fase (PM), con la onda modulada definida por

$$s(t) = A_c \cos\left[2\pi f_c t + k_p m(t)\right]$$

donde k_p es una constante y $m(t)$ es la señal de mensaje. El ruido aditivo $n(t)$ en la entrada del detector de fase es

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\,\text{sen}(2\pi f_c t)$$

Suponiendo que la relación portadora a ruido en la entrada del detector es alta comparada con la unidad. Determine *a*) la relación señal a ruido de salida, y *b*) la figura de mérito del sistema. Compare sus resultados con el sistema de FM para el caso de modulación senoidal.

2.55 Un sistema FDM utiliza la modulación de banda lateral única para combinar 12 señales de voz independientes y luego utiliza la modulación de frecuencia para transmitir la señal de banda base compuesta. Cada señal de voz tiene una potencia promedio P y ocupa la banda de frecuencia 0.3 a 3.4 kHz; el sistema le asigna un ancho de banda de 4 kHz. Para cada señal de voz, sólo se transmite la banda lateral inferior. Las ondas de la subportadora que se utilizan en la primera etapa de la modulación están definidas por

$$c_k(t) = A_k \cos(2\pi k f_0 t), \quad 0 \le k \le 11$$

La señal recibida consiste en una señal de FM transmitida más ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$.

a) Dibuje la densidad espectral de potencia de la señal que se produce a la salida del discriminador de frecuencia, indicando las componentes tanto de señal como de ruido.

b) Determine la relación entre las amplitudes A_k de la subportadora de manera que las señales de voz moduladas tengan relaciones de señal a ruido iguales.

2.56 En la explicación del efecto de umbral de FM que se presentó en la sección 2.13, describimos las condiciones de los clics que se hacen positivos y de los que se hacen negativos en términos de la envolvente $r(t)$ y la fase $\psi(t)$ del ruido de banda angosta $n(t)$. Reformule estas condiciones en términos de la componente en fase de $n_I(t)$ y de la componente en cuadratura $n_Q(t)$ de $n(t)$.

2.57 Mediante el uso del filtro de preénfasis que se muestra en la figura 2.50*a* y con una señal de voz como la onda moduladora, un transmisor de FM produce una señal que esencialmente se modula en frecuencia mediante frecuencias de audio inferiores y que se modula en fase a partir de frecuencias de audio superiores. Explique las razones de este fenómeno.

2.58 Suponga que las funciones de transferencia de los filtros de preénfasis y deénfasis de un sistema de FM se escalan de la manera siguiente:

$$H_{\text{pe}}(f) = k\left(1 + \frac{jf}{f_0}\right)$$

y

$$H_{\text{de}}(f) = \frac{1}{k}\left(\frac{1}{1 + jf/f_0}\right)$$

El factor de escalamiento k se elige de tal manera que la potencia promedio de la señal de mensaje enfatizada sea la misma que la de la señal de mensaje original $m(t)$.

a) Determine el valor de k que satisface este requerimiento para el caso en el que la densidad de potencia espectral de la señal de mensaje $m(t)$ es

$$S_M(f) = \begin{cases} \dfrac{S_0}{1 + (f/f_0)^2}, & -W \le f \le W \\ 0, & \text{en otro lado} \end{cases}$$

b) ¿Cuál es el valor correspondiente del factor de mejoramiento I producido al utilizar este par de filtros de preénfasis y deénfasis? Compare esta proporción con la que se obtiene en el ejemplo 2.6. El factor de mejora I se define por medio de la ecuación (2.160).

2.59 Un sistema de modulación de fase (PM) utiliza un par de filtros de preénfasis y deénfasis definidos por las funciones de transferencia

$$H_{pe}(f) = 1 + \frac{jf}{f_0}$$

y

$$H_{de}(f) = \frac{1}{1 + (jf/f_0)}$$

Demuestre que la mejora en la relación señal a ruido de salida producida al utilizar este par de filtros es

$$I = \frac{W/f_0}{\tan^{-1}(W/f_0)}$$

donde W es el ancho de banda del mensaje. Evalúe esta mejora para el caso en el que $W = 15$ kHz y $f_0 = 2.1$ kHz, y compare su resultado con el valor correspondiente para un sistema de FM.

Experimentos de computadora

2.60 En este experimento estudiamos el comportamiento del detector de envolvente que se muestra en la figura P2.5 para las siguientes especificaciones:

Resistencia de la fuente, $R_s = 75\ \Omega$
Resistencia de la carga, $R_l = 10\ \text{k}\Omega$
Capacitancia, $C = 0.01\ \mu\text{F}$

El diodo tiene una resistencia de 25 Ω cuando se polariza en directa y resistencia infinita cuando se polariza en inversa.

Calcule la forma de onda de la salida del detector de envolvente, suponiendo una onda de AM senoidal de entrada con 50% de modulación. La frecuencia de modulación es igual a 1 kHz y la frecuencia de la portadora corresponde a 20 kHz.

2.61 En este experimento continuamos el estudio de la malla de fase encadenada que se considera en la sección 2.14:

a) Calcule las variaciones en la frecuencia instantánea del oscilador controlado por voltaje en la malla para los siguientes parámetros de este último:

$$\text{Parámetro de la ganancia del lazo, } K_0 = \frac{50}{2\pi}\text{Hz}$$

$$\text{Frecuencia natural del lazo, } f_n = \frac{1}{2\pi}\text{Hz}$$

$$\text{Factor de amortiguamiento, } \zeta = 0.707$$

Realice los cálculos para los siguientes valores de paso de frecuencia: $\Delta f = 0.125,\ 0.5, \frac{7}{12}, \frac{2}{3}$ Hz.

b) Para los parámetros de la malla de fase encadenada que se especifica en el experimento 2 de la sección 2.14, calcule cómo afectan las variaciones en la desviación de frecuencia relativa $\Delta f \cdot f_m / f_n^2$ al error de fase pico de la malla de fase encadenada.

PROYECTO UNI / ASDI (SAREC) / FEC
UNIVERSIDAD NACIONAL DE INGENIERIA
FACULTAD DE ELECTROTECNIA Y COMPUTACION

CAPÍTULO 3

MODULACIÓN POR PULSOS

Este capítulo, que representa la transición de las comunicaciones analógicas a las digitales, comprende los siguientes temas:

- *Muestreo, el cual es básico para todas las formas de modulación por pulsos.*
- *Modulación por amplitud de pulsos, que es la forma más simple de modulación por pulsos.*
- *Cuantización, la cual, cuando se combina con el muestreo, permite la representación de una señal analógica en forma discreta tanto en amplitud como en tiempo.*
- *Modulación por codificación de pulsos, que es el método estándar para la transmisión de una señal de mensaje analógica por medios digitales.*
- *Multiplexado por división de tiempo, que permite compartir el tiempo de un canal común a una pluralidad de usuarios mediante la modulación por pulsos.*
- *Multiplexores digitales, que combinan muchos flujos de bits lentos en un solo flujo más rápido.*
- *Otras formas de modulación digital por pulsos, a saber, la modulación delta y modulación diferencial por codificación de pulsos.*
- *Predicción lineal, que es básica para la codificación de señales de mensaje analógicas a tasas de bits reducidas como en la modulación diferencial por codificación de pulsos.*
- *Formas adaptables de modulación diferencial por codificación de pulsos y modulación delta.*
- *El estándar de codificación de audio MPEG-1, que es un sistema de compresión perceptualmente sin pérdidas y transparente.*

3.1 *Introducción*

En la *modulación de onda continua (CW)*, que estudiamos en el capítulo 2, algunos parámetros de la onda portadora senoidal variaban continuamente en función de la señal del mensaje. Esto contrasta directamente con la modulación por pulsos, que estudiamos en el presente capítulo. En la *modulación por pulsos*, algún parámetro del tren de pulsos se varía de acuerdo con la señal del mensaje. Es posible distinguir dos familias de modulación por pulsos: *modulación analógica por pulsos* y *modulación digital por pulsos*. En la modulación analógica por pulsos se utiliza un tren de pulsos periódicos como la onda portadora y se varían algunos datos característicos de cada pulso (por ejemplo, amplitud, duración o posición) en forma continua de acuerdo con el correspondiente valor *de muestreo* de la señal del mensaje. Del mismo modo, en la modulación analógica por pulsos, la información se transmite básicamente en forma analógica, aunque la transmisión ocurre en tiempos discretos. En la modulación digital por pulsos, en cambio, la señal del mensaje se representa en una forma que es discreta tanto en tiempo como en amplitud, permitiendo de ese modo su transmisión en forma digital como una secuencia de *pulsos codificados*; esta forma de transmisión de señales *no* tiene contraparte de onda continua.

El uso de pulsos codificados para la transmisión de señales analógicas portadoras de información representa un ingrediente básico en la aplicación de las comunicaciones digitales. En consecuencia, es posible que este capítulo se considere como una transición de las comunicaciones analógicas a las digitales en nuestro estudio de los principios de los sistemas de comunicación. Iniciamos la exposición describiendo el proceso de muestreo, que es fundamental para todos los sistemas de modulación por pulsos, ya sea analógica o digital.

3.2 *Proceso de muestreo*

El *proceso de muestreo* suele describirse en el dominio del tiempo. En consecuencia, es una operación que resulta básica para el procesamiento de señales y comunicaciones digitales. Mediante el uso del proceso de muestreo, una señal analógica se convierte en una secuencia correspondiente de muestras que suelen espaciarse uniformemente en el tiempo. Sin duda, para que un procedimiento de este tipo tenga utilidad práctica, resulta necesario que elijamos adecuadamente la frecuencia de muestreo, de manera que la secuencia de muestras defina en forma única la señal analógica original. Ésta es la esencia del teorema del muestreo, el cual se deduce en lo que sigue.

Considere una señal arbitraria $g(t)$ de energía finita, que se especifica para todo tiempo. Un segmento de la señal $g(t)$ se muestra en la figura 3.1*a*. Suponga que muestreamos la señal $g(t)$ instantáneamente y a una velocidad uniforme, una vez cada T_s segundos. En consecuencia, obtenemos una secuencia infinita de muestras separadas por T_s segundos y denotadas por $\{g(nT_s)\}$, donde n toma todos los posibles valores enteros. Nos referimos a T_s como el *periodo de muestreo*, y a su recíproco $f_s = 1/T_s$ como la *frecuencia de muestreo*. Esta forma ideal de muestreo recibe el nombre de *muestreo instantáneo*.

Sea que $g_\delta(t)$ denote la señal que se obtiene al ponderar individualmente los elementos de una secuencia periódica de funciones delta separadas por T_s segundos mediante una secuencia de números $\{g(nT_s)\}$, como se indica mediante (vea la figura 3.1*b*)

$$g_\delta(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\,\delta(t - nT_s) \tag{3.1}$$

Nos referimos a $g_\delta(t)$ como la *señal muestreada ideal*. El término $\delta(t - nT_s)$ representa una función delta ubicada en el tiempo $t = nT_s$. De la definición de la función delta, recordamos que tal función idealizada tiene un área unitaria; vea el apéndice 2. De ese modo, podemos considerar al factor que multiplica $g(nT_s)$ en la ecuación (3.1) como una "masa" asignada a la función delta $\delta(t - nT_s)$. Una función delta ponderada de esta manera se aproxima bastante bien por medio de

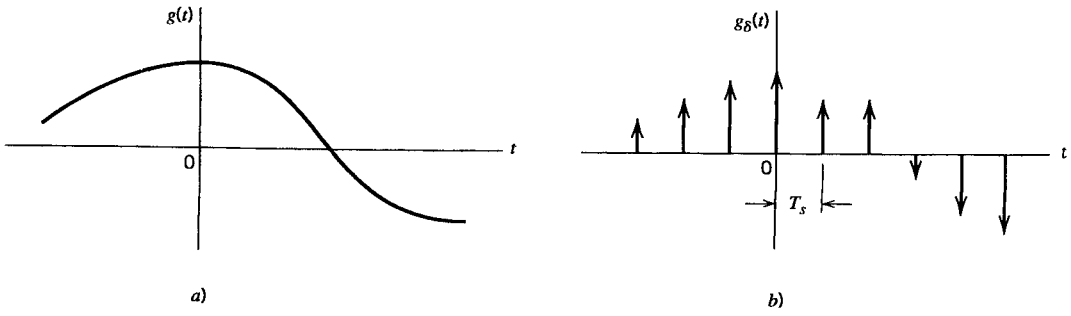


FIGURA 3.1 El proceso de muestreo. *a*) Señal analógica. *b*) Versión muestreada instantáneamente de la señal analógica.

un pulso rectangular de duración Δt y amplitud $g(nT_s)/\Delta t$; *cuanto más pequeña hagamos* Δt, *tanto mejor será la aproximación.*

Utilizando la tabla de pares de transformadas de Fourier, podemos escribir (vea la última línea de fórmulas de la tabla A6.3)

$$g_\delta(t) \rightleftharpoons f_s \sum_{m=-\infty}^{\infty} G(f - mf_s) \tag{3.2}$$

donde $G(f)$ es la transformada de Fourier de la señal original $g(t)$, y f_s es la frecuencia de muestreo. La ecuación (3.2) establece que *el proceso de muestrear uniformemente una señal en tiempo continuo de energía finita produce un espectro periódico con un periodo igual a la frecuencia de muestreo.*

Otra expresión útil para la transformada de Fourier de una señal muestreada $g_\delta(t)$ se puede obtener al tomar la transformada de Fourier de ambos miembros de la ecuación (3.1) y observar que la transformada de Fourier de la función delta $\delta(t - nT_s)$ es igual a $\exp(-j2\pi nfT_s)$. Dejemos que $G_\delta(f)$ denote la transformada de Fourier de $g_\delta(t)$. Por tanto, es posible escribir

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g(nT_s) \exp(-j2\pi nfT_s) \tag{3.3}$$

Esta relación se conoce como la *transformada de Fourier en tiempo discreto.* Puede considerarse como una representación compleja de la serie de Fourier de la función de frecuencia periódica $G_\delta(f)$, definiendo la secuencia de muestras $\{g(nT_s)\}$ a los coeficientes de la expansión.

Las relaciones, de acuerdo a como se obtienen aquí, se aplican a cualquier señal en tiempo continuo $g(t)$ de energía finita y duración infinita. Sin embargo, suponga que la señal $g(t)$ es *estrictamente limitada en banda,* sin componentes de frecuencia mayores que W Hertz. Es decir, la transformada de Fourier $G(f)$ de la señal $g(t)$ tiene la propiedad de que $G(f)$ es cero para $|f| \geq W$, como se ilustra en la figura 3.2*a*; la forma del espectro que se muestra en la figura se escogió exclusivamente con fines ilustrativos. Suponga también que elegimos el periodo de muestreo $T_s = 1/2W$. En ese caso el espectro correspondiente $G_\delta(f)$ de la señal muestreada $g_\delta(t)$ es como se muestra en la figura 3.2*b*. Al hacer $T_s = 1/2W$ en la ecuación (3.3), se obtiene

$$G_\delta(f) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi nf}{W}\right) \tag{3.4}$$

De acuerdo con la ecuación (3.2), vemos de inmediato que la transformada de Fourier $g_\delta(t)$ también puede expresarse como

$$G_\delta(f) = f_s G(f) + f_s \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} G(f - mf_s) \tag{3.5}$$

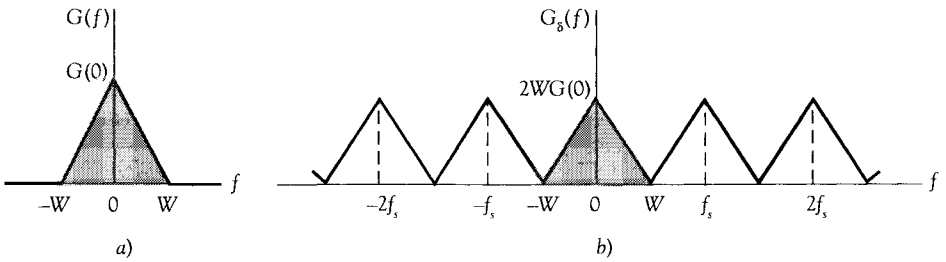


FIGURA 3.2 *a*) Espectro de una señal estrictamente limitada en banda $g(t)$. *b*) Espectro de la versión muestreada de $g(t)$ para un periodo de muestreo $T_s = 1/2W$.

Por tanto, bajo las siguientes dos condiciones:

1. $G(f) = 0$ para $|f| \geq W$
2. $f_s = 2W$

encontramos a partir de la ecuación (3.5) que

$$G(f) = \frac{1}{2W} G_\delta(f), \quad -W < f < W \tag{3.6}$$

Al sustituir la ecuación (3.4) en la (3.6), también es posible que escribamos

$$G(f) = \frac{1}{2W} \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi n f}{W}\right), \quad -W < f < W \tag{3.7}$$

Por tanto, si los valores de muestreo $g(n/2W)$ de una señal $g(t)$ se especifican para toda n, entonces la transformada de Fourier $G(f)$ de la señal está determinada en forma única al utilizar la transformada de Fourier en tiempo discreto de la ecuación (3.7). Debido a que $g(t)$ se relaciona con $G(f)$ por medio de la transformada inversa de Fourier, se concluye que la misma señal $g(t)$ se determina en forma única mediante valores de muestreo $g(n/2W)$ para $-\infty < n < \infty$. En otras palabras, la secuencia $\{g(n/2W)\}$ tiene toda la información contenida en $g(t)$.

Considere a continuación el problema de la reconstrucción de la señal $g(t)$ a partir de la secuencia de los valores muestreados $\{g(n/2W)\}$. Al sustituir la ecuación (3.7) en la fórmula para la transformada inversa de Fourier que define a $g(t)$ en términos de $G(f)$, obtenemos

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty} G(f) \exp(j2\pi f t)\, df \\ &= \int_{-W}^{W} \frac{1}{2W} \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \exp\left(-\frac{j\pi n f}{W}\right) \exp(j2\pi f t)\, df \end{aligned}$$

Al intercambiar el orden de la sumatoria y de la integral:

$$g(t) = \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \frac{1}{2W} \int_{-W}^{W} \exp\left[j2\pi f\left(t - \frac{n}{2W}\right)\right] df \tag{3.8}$$

El término integral en la ecuación (3.8) se evalúa fácilmente, produciéndose el resultado final

$$\begin{aligned} g(t) &= \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \frac{\operatorname{sen}(2\pi W t - n\pi)}{(2\pi W t - n\pi)} \\ &= \sum_{n=-\infty}^{\infty} g\left(\frac{n}{2W}\right) \operatorname{senc}(2W t - n), \quad -\infty < t < \infty \end{aligned} \tag{3.9}$$

La ecuación (3.9) proporciona una *fórmula de interpolación* para reconstruir la señal $g(t)$ a partir de la secuencia de valores muestreados $\{g(n/2W)\}$, con la función $\operatorname{senc}(2Wt)$ desempeñando el papel de una *función de interpolación*. Cada muestra se mutiplica por una versión atrasada de la función de interpolación, y todas las formas de onda que resultan se suman para obtener $g(t)$.

Ahora podemos establecer el *teorema de muestreo* para señales estrictamente limitadas en banda de energía finita en dos partes equivalentes, las cuales aplicamos, respectivamente, al transmisor y al receptor del sistema de modulación por pulsos:

1. *Una señal limitada en banda de energía finita, que no tiene componentes de frecuencia mayores que W hertz, se describe por completo especificando los valores de la señal en instantes separados por 1/2W segundos.*
2. *Una señal limitada en banda de energía finita, que no tiene componentes de frecuencia mayores que W hertz, puede recuperarse completamente a partir del conocimiento de sus muestras tomadas a una velocidad de 2W muestras por segundo.*

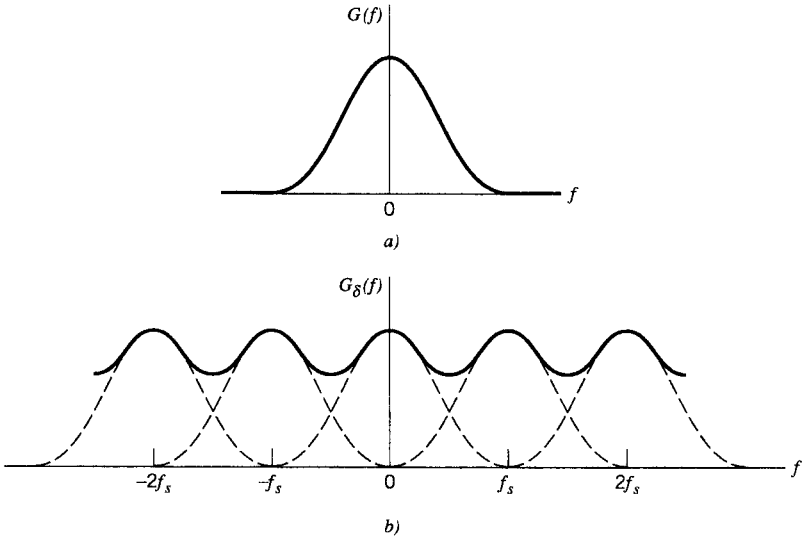


FIGURA 3.3 *a*) Espectro de una señal. *b*) Espectro de una versión submuestreada de la señal presentando el fenómeno de traslape.

La frecuencia de muestreo de $2W$ muestras por segundo, para un ancho de banda de señal de W hertz, se denomina la *frecuencia de Nyquist;* su recíproco $1/2W$ (medido en segundos) se denomina el *intervalo de Nyquist.*

La deducción del teorema de muestreo, según se describe aquí, se basa en la suposición de que la señal $g(t)$ es estrictamente limitada en banda. Sin embargo, en la práctica una señal que porta información *no* es estrictamente limitada en banda, con el resultado de que se encuentra cierto grado de submuestreo. En consecuencia, algo de *traslape* se produce mediante el proceso de muestreo. El traslape se refiere al fenómeno de que una componente de alta frecuencia en el espectro de la señal adquiere aparentemente la identidad de una frecuencia inferior en el espectro de su versión muestreada, como se ilustra en la figura 3.3. El espectro de traslape, mostrado por medio de una curva continua en la figura 3.3*b*, pertenece a una versión "submuestreada" de la señal del mensaje representada por el espectro de la figura 3.3*a*.

Para combatir los efectos del traslape en la práctica, es posible utilizar dos medidas correctivas, como se describe a continuación:

1. Antes del muestreo se emplea un *filtro* pasobajas *de antitraslape* para atenuar aquellas componentes de alta frecuencia de la señal que no son esenciales para la información que esta misma transporta.
2. La señal filtrada se muestrea a una frecuencia un poco mayor que la de Nyquist.

El uso de una frecuencia de muestreo mayor que la de Nyquist tiene también el efecto benéfico de facilitar el diseño del *filtro de reconstrucción* que se utiliza para recuperar la señal original de su versión muestreada. Considere el ejemplo de una señal de mensaje sobre la que se ha realizado filtrado antitraslape (pasobajas), produciendo el espectro que se muestra en la figura 3.4*a*. El espectro correspondiente de la versión muestreada instantáneamente de la señal se ilustra en la figura 3.4*b*, considerando una frecuencia de muestreo superior a la de Nyquist. De acuerdo con la figura 3.4*b*, vemos de inmediato que el diseño del filtro de reconstrucción puede especificarse del modo siguiente (vea la figura 3.4*c*):

- El filtro de reconstrucción es pasobajas con una banda de paso que se extiende de $-W$ a W, el cual se determina por medio de un filtro antitraslape.
- El filtro tiene una banda de transición que se extiende (para frecuencias positivas) desde W hasta $f_s - W$, donde f_s es la frecuencia de muestreo.

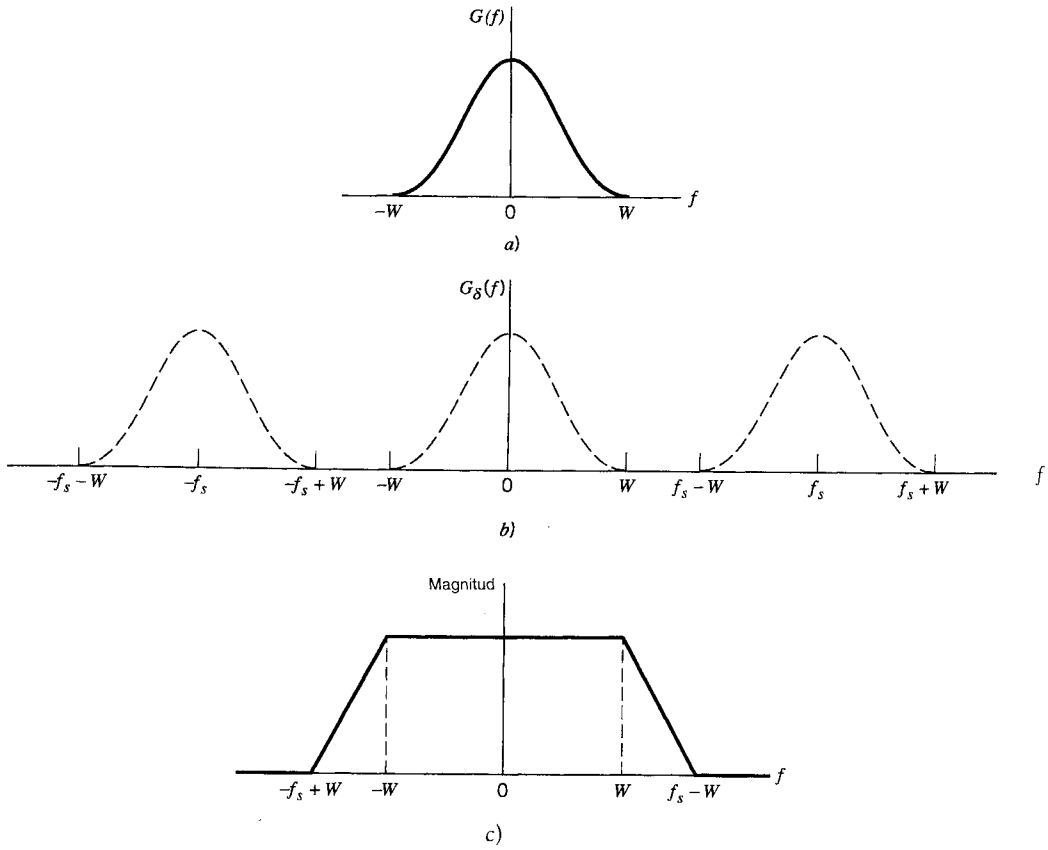


FIGURA 3.4 *a*) Espectro filtrado antitraslape de una señal que contiene información. *b*) Espectro de una versión muestreada instantáneamente de la señal, suponiendo el uso de una frecuencia de muestreo mayor que la frecuencia de Nyquist. *c*) Respuesta en magnitud del filtro de reconstrucción.

El hecho de que el filtro de reconstrucción cuente con una banda de transición bien definida significa que éste es físicamente realizable.

3.3 *Modulación por amplitud de pulsos*

Ahora que comprendemos la esencia del proceso de muestreo, estamos listos para definir formalmente la modulación por amplitud de pulsos, la cual es la forma más simple y fundamental de la modulación analógica por pulsos. En la *modulación por amplitud de pulsos (PAM), las amplitudes de pulsos espaciados regularmente se varían en proporción a los valores de muestreo correspondientes de una señal de mensaje continua;* los pulsos pueden ser de forma rectangular o de alguna otra forma apropiada. La modulación por amplitud de pulsos, según se define aquí, es un poco similar al muestreo natural, donde la señal del mensaje se multiplica por un tren periódico de pulsos rectangulares. Sin embargo, en el muestreo natural la parte superior de cada pulso rectangular modulado varía con la señal del mensaje, en tanto que en la señal de PAM esa parte se mantiene plana; el muestreo natural se investiga un poco más en el problema 3.2.

La forma de onda de una señal de PAM se ilustra en la figura 3.5. La curva punteada en esta figura delinea la forma de onda de una señal de mensaje $m(t)$, y la secuencia de pulsos rectangula-

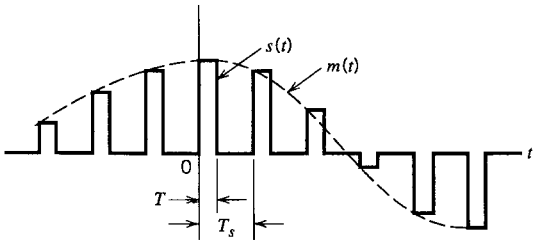


FIGURA 3.5 Muestras de cresta plana, representando una señal analógica.

res modulados en amplitud e ilustrados como líneas continuas representa a las señales $s(t)$ de la PAM correspondientes. Hay dos operaciones que incluyen la generación de una señal de PAM:

1. *Muestreo instantáneo* de la señal de mensaje $m(t)$ cada T_s segundos, donde la frecuencia de muestreo $f_s = 1/T_s$ se elige de acuerdo con el teorema de muestreo.
2. *Alargamiento* de la duración de cada muestra así obtenida hasta cierto valor constante T.

En la tecnología de circuitos digitales se hace referencia en forma conjunta a estas dos operaciones como "muestreo y retención". Una razón importante del alargamiento intencional de la duración de cada muestra es evitar el uso de un ancho de banda de canal excesivo, ya que el ancho de banda es inversamente proporcional a la duración del pulso. Sin embargo, es necesario tener cuidado con respecto a qué tan larga se hace la duración de muestreo T, según revela el siguiente análisis.

Sea $s(t)$ la secuencia de pulsos de cresta plana generados de la manera que se describe en la figura 3.5. Podemos expresar la señal PAM como

$$s(t) = \sum_{n=-\infty}^{\infty} m(nT_s)h(t - nT_s) \tag{3.10}$$

donde T_s es el *periodo de muestreo* y $m(nT_s)$ es el valor de la muestra de $m(t)$ obtenido en el tiempo $t = nT_s$. La $h(t)$ es un pulso rectangular estándar de amplitud unitaria y duración T que se define del modo siguiente (vea la figura 3.6*a*):

$$h(t) = \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t = 0, t = T \\ 0, & \text{en otro caso} \end{cases} \tag{3.11}$$

Por definición, la versión de $m(t)$ muestreada instantáneamente está dada por

$$m_\delta(t) = \sum_{n=-\infty}^{\infty} m(nT_s)\ \delta(t - nT_s) \tag{3.12}$$

donde $\delta(t - nT_s)$ es una función delta recorrida en el tiempo. Por tanto, convolucionando $m_\delta(t)$ con el pulso $h(t)$, obtenemos

$$\begin{aligned} m_\delta(t) \bigstar h(t) &= \int_{-\infty}^{\infty} m_\delta(\tau)h(t - \tau)\ d\tau \\ &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} m(nT_s)\ \delta(\tau - nT_s)h(t - \tau)\ d\tau \\ &= \sum_{n=-\infty}^{\infty} m(nT_s) \int_{-\infty}^{\infty} \delta(\tau - nT_s)h(t - \tau)\ d\tau \end{aligned} \tag{3.13}$$

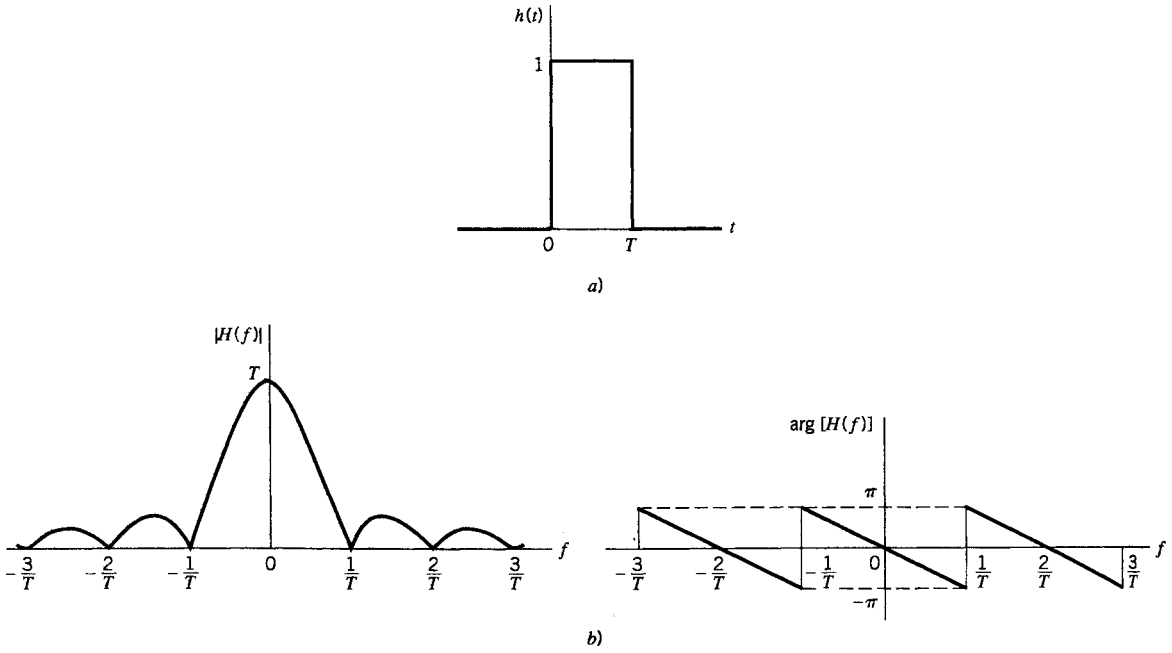


FIGURA 3.6 *a*) Pulso rectangular $h(t)$. *b*) Espectro $H(f)$, integrado por la magnitud $|H(f)|$ y la fase $\arg[H(f)]$.

Utilizando la propiedad de filtrado de la función delta (vea el apéndice 2), obtenemos

$$m_\delta(t) \bigstar h(t) = \sum_{n=-\infty}^{\infty} m(nT_s)h(t - nT_s) \tag{3.14}$$

De acuerdo con las ecuaciones (3.10) y (3.14) se concluye que la señal $s(t)$ de PAM es matemáticamente equivalente a la convolución de $m_\delta(t)$, la versión de $m(t)$ muestreada instantáneamente, y el pulso $h(t)$, según indica

$$s(t) = m_\delta(t) \bigstar h(t) \tag{3.15}$$

Tomando la transformada de Fourier de ambos miembros de la ecuación (3.15) y reconociendo que la convolución de dos funciones del tiempo se transforma en una multiplicación de sus transformadas de Fourier respectivas, tenemos

$$S(f) = M_\delta(f)H(f) \tag{3.16}$$

donde $S(f) = F[s(t)]$, $M_\delta(f) = F[m_\delta(t)]$ y $H(f) = F[h(t)]$. Adaptando la ecuación (3.2) al problema presente, advertimos que la transformada de Fourier $M_\delta(f)$ se relaciona con la transformada de Fourier $M(f)$ de la señal de mensaje original $m(t)$ como sigue

$$M_\delta(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s) \tag{3.17}$$

donde f_s es la frecuencia de muestreo. Por tanto, la sustitución de la ecuación (3.17) en la (3.16) produce

$$S(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s)H(f) \tag{3.18}$$

Dada una señal $s(t)$ de PAM cuya transformada de Fourier $S(f)$ se define como la ecuación (3.18), ¿cómo recuperamos la señal de mensaje original $m(t)$? Como paso inicial en esta

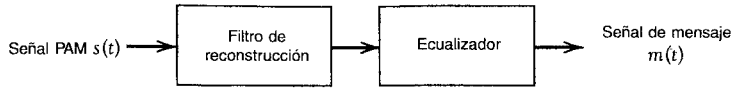


FIGURA 3.7 Sistema para recuperar la señal de mensaje $m(t)$ de la señal PAM $s(t)$.

reconstrucción, es posible pasar $s(t)$ por un filtro pasobajas cuya respuesta en frecuencia se define en la figura 3.4*c*. Aquí se supone que el mensaje está limitado al ancho de banda W y que la frecuencia de muestreo f_s es mayor que la de Nyquist $2W$. En ese caso, a partir de la ecuación (3.18) encontramos que el espectro que resulta del filtro es igual a $M(f)H(f)$. Esta salida es equivalente a pasar la señal de mensaje original $m(t)$ por otro filtro pasobajas de respuesta en frecuencia $H(f)$.

De acuerdo con la ecuación (3.11), advertimos que la transformada de Fourier del pulso $h(t)$ está dada por

$$H(f) = T\,\text{senc}(fT)\,\exp(-j\pi fT) \tag{3.19}$$

misma que se grafica en la figura 3.6*b*. Por tanto, vemos que al utilizar muestras de cresta plana para generar una señal PAM hemos introducido *distorsión de amplitud*, así como un *retardo* de $T/2$. Este efecto es bastante similar a la variación en la transmisión con la frecuencia que provoca el tamaño finito de la abertura de la exploración óptica en la televisión. En consecuencia, la distorsión que provoca el empleo de la modulación de amplitud por pulsos para transmitir una señal analógica que contiene información se conoce como el *efecto de apertura*.

Es posible corregir esta distorsión conectando un *ecualizador* en cascada con el filtro de reconstrucción pasobajas, como se ilustra en la figura 3.7. El ecualizador tiene el efecto de reducir la pérdida en banda del filtro de reconstrucción cuando la frecuencia aumenta de tal manera que compensa el efecto de apertura. Idealmente, la respuesta en magnitud está dada por

$$\frac{1}{|H(f)|} = \frac{1}{T\,\text{senc}(fT)} = \frac{\pi f}{\text{sen}(\pi fT)} \tag{3.20}$$

La cantidad de ecualización que se necesita en la práctica casi siempre es pequeña. En realidad, para un ciclo de trabajo $T/T_s \leq 0.1$, la distorsión de amplitud es menor que 0.5%, en cuyo caso la necesidad de ecualización quizá se omita completamente.

La transmisión de una señal PAM impone requerimientos bastante estrictos en las respuestas en magnitud y fase del canal, debido a la duración relativamente corta de los pulsos transmitidos. Además, el desempeño ante el ruido de un sistema PAM nunca puede ser mejor que la transmisión de la señal de la banda base. En consecuencia, encontramos que en la transmisión a largas distancias, la PAM se usaría sólo como un medio de procesamiento de mensajes en el multiplexado por división de tiempo, a partir del cual se realiza la conversión a otra forma de modulación por pulsos; el multiplexado por división de tiempo se explica en la sección 3.9.

3.4 *Otras formas de modulación por pulsos*

En un sistema de modulación por pulsos podemos utilizar el ancho de banda incrementado que consumen los pulsos para mejorar el desempeño del sistema con respecto al ruido. Lo anterior puede conseguirse representando los valores de muestreo de las señales de mensaje mediante alguna propiedad del pulso aparte de la amplitud:

- *Modulación por duración de pulsos* (PDM), conocida también como *modulación por ancho de pulso*, donde las muestras de la señal del mensaje se utilizan para variar la duración de los pulsos individuales en la portadora.

- *Modulación por posición de pulso* (PPM), donde la posición de un pulso relativa a su tiempo de ocurrencia sin modulación se varía de acuerdo con la señal del mensaje.

Estas dos formas de modulación por pulsos se ilustran en la figura 3.8 para el caso de una onda moduladora senoidal.

En la PDM, los pulsos largos consumen bastante potencia mientras no lleven información adicional. Si esta potencia no utilizada se sustrae de la PDM de manera que únicamente se preserven las transiciones de tiempo, obtenemos la PPM. En consecuencia, esta última es una forma más eficiente de modulación por pulsos que la PDM.

Puesto que en un sistema PPM la información que se transmite se encuentra contenida en posiciones relativas de los pulsos modulados, la presencia de ruido aditivo afecta el desempeño de un sistema de este tipo, falseando el tiempo en el cual se estima que ocurren los pulsos modulados. Es posible establecer inmunidad al ruido haciendo que el pulso se construya tan rápidamente que resulte muy corto el intervalo de tiempo durante el cual el ruido puede ejercer cualquier perturbación. En realidad, el ruido aditivo no tendría ningún efecto en las posiciones de los pulsos si los pulsos recibidos fueran perfectamente rectangulares, ya que la presencia de ruido sólo introduce perturbaciones verticales. Sin embargo, la recepción de pulsos perfectamente rectangulares requeriría un ancho de banda de canal infinito, lo cual resulta impráctico. Por tanto, con un ancho de banda de canal finito en la práctica, encontramos que los pulsos recibidos tienen un tiempo de levantamiento finito, de modo que el desempeño del receptor PPM resulta afectado por el ruido, que es lo esperado.

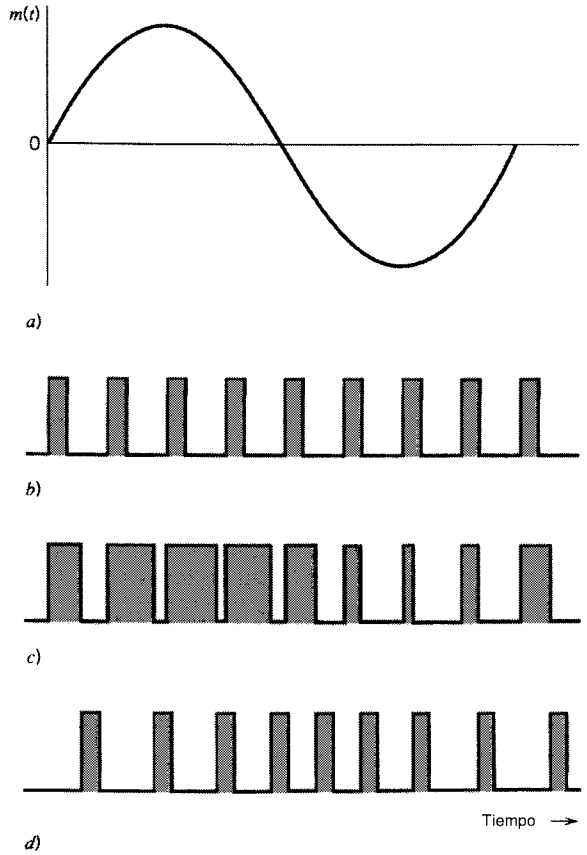


FIGURA 3.8 Ilustración de dos formas diferentes de modulación por tiempo de pulso para el caso de una onda moduladora senoidal. *a*) Onda moduladora. *b*) Portadora de pulso. c) Onda PDM. *d*) Onda PPM.

Como en un sistema de modulación de onda continua, el desempeño ante el ruido de un sistema PPM puede describirse en términos de la relación señal a ruido de salida (SNR). Además, para determinar el mejoramiento ante el ruido que produce la PPM sobre la transmisión de la banda base de una señal de mensaje, es factible utilizar la figura de mérito definida como la relación señal a ruido de salida del sistema PPM dividida por la relación señal a ruido del canal; vea la sección 2.10. Suponiendo que la potencia promedio del ruido del canal es pequeña en comparación con la potencia del pulso pico, la figura de mérito del sistema PPM es proporcional al cuadrado del ancho de banda de transmisión B_T normalizado con respecto al ancho de banda del mensaje W. Sin embargo, cuando la relación señal a ruido de entrada disminuye por debajo de un valor crítico, el sistema sufre una pérdida de la señal de mensaje deseada en la salida del receptor. Esto es, un sistema PPM sufre de su propio efecto de umbral.

3.5 *Compromiso ancho de banda-ruido*

En el contexto del desempeño ante el ruido, un sistema PPM constituye la forma óptima de modulación analógica por pulsos. El análisis de ruido de un sistema PPM revela que los sistemas de modulación por posición de pulsos (PPM) y los de modulación de frecuencia (FM) exhiben un desempeño ante el ruido similar, como se resume aquí.[1]

1. Ambos sistemas tienen una figura de mérito proporcional al cuadrado del ancho de banda de transmisión normalizado con respecto al ancho de banda del mensaje.
2. Ambos sistemas presentan un efecto de umbral cuando se reduce la relación señal a ruido.

La aplicación práctica del punto 1 es que, en términos de un compromiso de ancho de banda incrementado por un desempeño mejorado ante el ruido, lo mejor que podemos conseguir con los sistemas de modulación de onda continua y de modulación analógica por pulsos es seguir una *ley al cuadrado*. La pregunta que surge en este momento en la exposición es la siguiente: ¿resulta posible obtener un compromiso mejor que una ley al cuadrado? La respuesta es un sí definitivo, y la *modulación digital por pulsos* es la forma de lograrlo. El uso de un método de este tipo constituye una desviación radical de la modulación de onda continua.

De manera específica, en una forma básica de la modulación digital por pulsos que se conoce como *modulación por codificación de pulsos* (PCM),[2] una señal del mensaje se representa en forma discreta tanto en tiempo como en amplitud. Esta forma de representación de la señal permite la transmisión de la señal del mensaje como una secuencia de *pulsos binarios codificados*. Dada una secuencia de este tipo, el efecto del ruido del canal en la salida del receptor puede reducirse hasta un nivel ignorable haciendo simplemente que la potencia promedio de la onda binaria transmitida PCM sea lo suficientemente grande en comparación con la correspondiente al ruido.

Dos procesos fundamentales están implicados en la generación de una onda PCM binaria: *muestreo* y *cuantización*. El proceso de muestreo se ocupa de la representación en tiempo discreto de la señal del mensaje. Para su aplicación apropiada, tenemos que seguir el teorema de muestreo que se describió en la sección 3.2. El proceso de cuantización tiene que ver con la representación discreta de la amplitud de la señal del mensaje; la cuantización es un proceso nuevo, cuyos detalles se describen en la sección siguiente. Por ahora basta señalar que el uso combinado del muestreo y la cuantización permite la transmisión de una señal de mensaje en forma codificada que, a su vez, posibilita realizar una *ley exponencial* para el compromiso ancho de banda-ruido, lo cual se demuestra también en la sección que sigue.

3.6 *Proceso de cuantización*[3]

Una señal continua, tal como la voz, tiene un intervalo continuo de amplitudes y, por tanto, sus muestras tienen un intervalo continuo de amplitud. En otras palabras, dentro del intervalo de

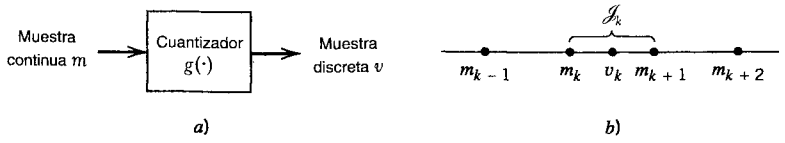


FIGURA 3.9 Descripción de un cuantizador sin memoria.

amplitud finito de la señal, encontramos un número infinito de niveles de amplitud. De hecho, no es necesario transmitir las amplitudes exactas de las muestras. Cualquier sentido humano (el oído o el ojo), como un receptor final, puede detectar únicamente diferencias de intensidad finitas, lo cual quiere decir que la señal continua original *puede aproximarse* mediante una señal construida a partir de amplitudes discretas elegidas de un conjunto disponible sobre una base de error mínimo. La existencia de un número finito de niveles de amplitud discreta es una condición básica de la modulación por codificación de pulsos. Si asignamos con claridad los niveles de amplitud discretos con un espaciamiento bastante próximo, es posible lograr que la señal aproximada prácticamente no se distinga de la señal continua original.

La *cuantización* de amplitud se define como *el proceso de transformar la muestra de amplitud* $m(nT_s)$ *de una señal de mensaje* $m(t)$ *en el tiempo* $t = nT_s$ *en una amplitud discreta de* $v(nT_s)$ *tomada de un conjunto finito de amplitudes posibles.* Suponemos que el proceso de cuantización *no tiene memoria* y es *instantáneo,* lo que equivale a que la transformación en el tiempo $t = nT_s$ no resulta afectada por muestras anteriores o posteriores de la señal del mensaje. Esta forma simple de cuantización escalar, aunque no óptima, se usa mucho en la práctica.

Cuando se trata con un cuantizador sin memoria, podemos simplificar la notación eliminando el índice de tiempo. En consecuencia, es posible utilizar el símbolo m en lugar de $m(nT_s)$, como se indica en el diagrama de bloques de un cuantizador que se muestra en la figura 3.9*a*. En esas condiciones, como se muestra en la figura 3.9*b*, la amplitud de la señal m se especifica por medio del índice k si se ubica dentro de la *celda de partición*

$$\mathcal{J}_k: \{m_k < m \leq m_{k+1}\}, \qquad k = 1, 2, \ldots, L \tag{3.21}$$

donde L es el número total de niveles de amplitud utilizados en el cuantizador. Las amplitudes discretas, m_k, $k = 1, 2, \ldots, L$, en la entrada del cuantizador reciben el nombre de *niveles de decisión* o *umbrales de decisión.* En la salida del cuantizador, el índice k se transforma en una amplitud v_k que representa todas las amplitudes de la celda $\mathcal{J}_k$; las amplitudes discretas v_k, $k = 1, 2, \ldots, L$, reciben el nombre de *niveles de representación* o *de reconstrucción,* y el espaciamiento entre dos niveles de representación adyacentes se denomina *cuanto* o *tamaño de paso.* De esa manera, la salida del cuantizador v es igual a v_k si la muestra de la señal de entrada m pertenece al intervalo $\mathcal{J}_k$. La correspondencia (vea la figura 3.9*a*)

$$v = g(m) \tag{3.22}$$

es la *característica del cuantizador* que por definición es una función de escalera.

Los cuantizadores pueden ser de tipo *uniforme* o *no uniforme.* En el primer caso, los niveles de representación están uniformemente espaciados. En el otro caso, el cuantizador es no uniforme. En esta sección consideramos únicamente cuantizadores uniformes; los no uniformes se consideran en la sección 3.7. Las características del cuantizador también pueden ser del tipo *medio escalón* o *medio aumento.* La figura 3.10*a* muestra la característica de entrada-salida de un cuantizador uniforme de tipo medio escalón, el cual se denomina así porque el origen se encuentra a la mitad de un escalón de la gráfica de escalera. La figura 3.10*b* presenta la correspondiente a la característica de entrada-salida de un cuantizador uniforme de medio aumento, en el cual el origen se encuentra a la mitad de la parte ascendente de la gráfica similar a una escalera. Advierta que tanto el tipo de medio escalón como el de medio aumento de los cuantizadores uniformes que se ilustran en la figura 3.10 son *simétricos* alrededor del origen.

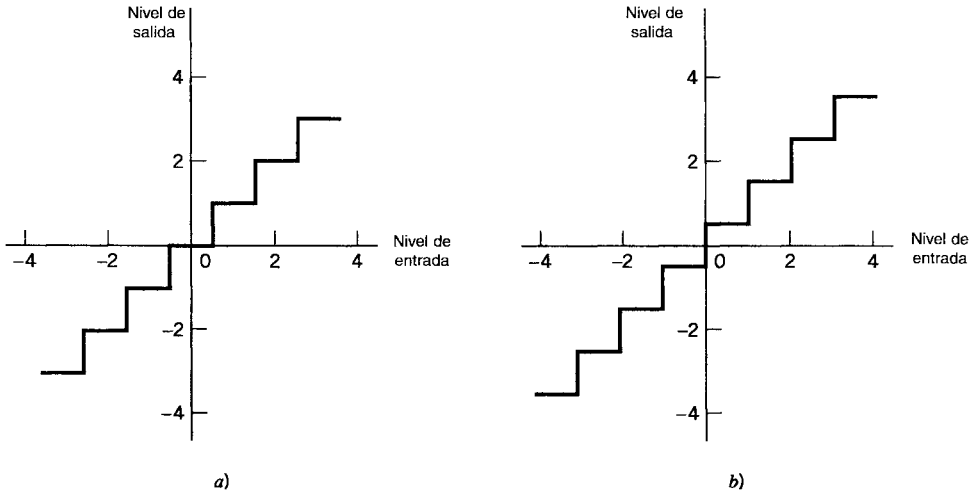


FIGURA 3.10 Dos tipos de cuantización: *a*) de medio escalón y *b*) de medio aumento.

■ RUIDO DE CUANTIZACIÓN

El uso de la cuantización introduce un error que se define como la diferencia entre la señal de entrada m y la señal de salida v. El error recibe el nombre de *ruido de cuantización*. La figura 3.11 ejemplifica una variación típica del ruido de cuantización como función del tiempo, suponiendo el uso de un cuantizador uniforme del tipo de medio escalón.

Dejemos que la entrada del cuantizador m sea el valor de la muestra de la variable aleatoria M de media cero. (Si la entrada tiene una media distinta de cero, siempre podemos eliminarla restando la media de la entrada y sumándola de nuevo después de la cuantización.) Un

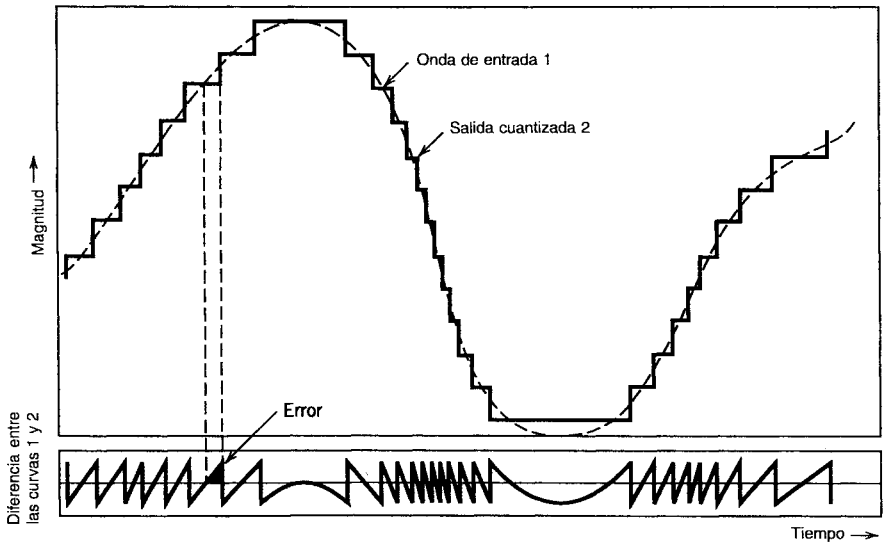


FIGURA 3.11 Ilustración del proceso de cuantización. (Adaptado de Bennett, con autorización de AT&T.)

cuantizador $g(\cdot)$ hace corresponder la variable aleatoria de entrada M de amplitud continua con una variable aleatoria discreta V; sus valores de las muestras respectivas m y v se relacionan por medio de la ecuación (3.22). Denotemos el error de cuantización por la variable aleatoria Q del valor de muestreo q. De tal modo, describimos

$$q = m - v \tag{3.23}$$

o, en forma correspondiente,

$$Q = M - V \tag{3.24}$$

Teniendo la entrada M media igual a cero y suponiendo que el cuantizador es simétrico como en la figura 3.10, se concluye que la salida del cuantizador V y, en consecuencia, el error de cuantización Q también tendrán media cero. De tal modo, en una caracterización estadística parcial del cuantizador en términos de la relación señal a ruido de salida (cuantización) sólo necesitamos determinar el valor cuadrático medio del error de cuantización Q.

Considere entonces la entrada m de amplitud continua en el intervalo $(-m_{\text{máx}}, m_{\text{máx}})$. Suponiendo un cuantizador uniforme del tipo de medio aumento que se ilustra en la figura 3.10*b*, encontramos que el tamaño del escalón del cuantizador está dado por

$$\Delta = \frac{2m_{\text{máx}}}{L} \tag{3.25}$$

donde L es el número total de niveles de representación. Para un cuantizador uniforme, el error de cuantización Q tendrá sus valores de las muestras delimitadas por $-\Delta/2 \leq q \leq \Delta/2$. Si el tamaño del escalón Δ es suficientemente pequeño (es decir, el número de niveles de representación L es lo bastante grande), resulta razonable suponer que el error de cuantización Q es una variable aleatoria *uniformemente distribuida*, y que el efecto de interferencia del ruido de cuantización en la entrada del cuantizador es similar al del ruido térmico. De tal modo, podemos expresar la función de densidad de probabilidad del error de cuantización Q del modo siguiente:

$$f_Q(q) = \begin{cases} \dfrac{1}{\Delta}, & -\dfrac{\Delta}{2} < q \leq \dfrac{\Delta}{2} \\ 0, & \text{en otro caso} \end{cases} \tag{3.26}$$

Sin embargo, para que esto sea cierto necesitamos asegurar que la señal de entrada *no* sobrecargue al cuantizador. En ese caso, con la media del error de cuantización igual a cero, su varianza σ_Q^2 es la misma que la del valor cuadrático medio:

$$\begin{aligned} \sigma_Q^2 &= E[Q^2] \\ &= \int_{-\Delta/2}^{\Delta/2} q^2 f_Q(q)\, dq \end{aligned} \tag{3.27}$$

Al sustituir la ecuación (3.26) en la (3.27), obtenemos

$$\begin{aligned} \sigma_Q^2 &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q^2\, dq \\ &= \frac{\Delta^2}{12} \end{aligned} \tag{3.28}$$

Por lo común, el número L-ario k, que denota el k-ésimo nivel de representación del cuantizador, se transmite al receptor en forma binaria. Denotemos mediante R el número de *bits por muestra* utilizados en la construcción del código binario. Entonces es posible escribir

$$L = 2^R \tag{3.29}$$

o, de modo equivalente,

$$R = \log_2 L \tag{3.30}$$

Por consiguiente, sustituyendo la ecuación (3.29) en la (3.25), obtenemos el tamaño de escalón

$$\Delta = \frac{2m_{\text{máx}}}{2^R} \tag{3.31}$$

El uso de la ecuación (3.31) en la (3.28) produce consecuentemente

$$\sigma_Q^2 = \tfrac{1}{3} m_{\text{máx}}^2 2^{-2R} \tag{3.32}$$

Sea P la potencia promedio de la señal de mensaje $m(t)$. Es posible expresar entonces la *relación señal a ruido de salida* de un cuantizador uniforme como

$$(SNR)_O = \frac{P}{\sigma_Q^2} \tag{3.33}$$

$$= \left(\frac{3P}{m_{\text{máx}}^2} \right) 2^{2R}$$

La ecuación (3.33) muestra que la relación señal a ruido de salida del cuantizador aumenta *exponencialmente* con el número creciente de bits por muestra, R. Reconociendo que un aumento en R requiere un incremento proporcional en el ancho de banda (transmisión) del canal B_T, vemos consecuentemente que el uso de un código binario para la representación de una señal de mensaje (como en la modulación por codificación de pulsos) ofrece un método más eficiente que cualquier modulación de frecuencia (FM) o modulación por posición de pulsos (PPM) para el compromiso entre el mayor ancho de banda del canal y el mejor desempeño ante el ruido. Al plantear este enunciado, supusimos que los sistemas de FM y PPM están limitados por el ruido del receptor, en tanto que el sistema de modulación por codificación binaria está limitado por el ruido de cuantización. Ampliaremos este tema al final de la sección 3.8.

Ejemplo 3.1 Señal moduladora senoidal

Considere el caso especial de una señal moduladora senoidal de carga completa y amplitud A_m, que utiliza todos los niveles de representación que se proporcionan. La potencia promedio de la señal es (suponiendo una carga de 1 ohm)

$$P = \frac{A_m^2}{2}$$

El intervalo total de la entrada del cuantizador es $2A_m$, debido a que la señal moduladora oscila entre $-A_m$ y A_m. Por tanto, es posible fijar $m_{\text{máx}} = A_m$, en cuyo caso el uso de la ecuación (3.32) produce la potencia promedio (varianza) del ruido de cuantización como

$$\sigma_Q^2 = \tfrac{1}{3} A_m^2 2^{-2R}$$

De tal manera, la relación señal a ruido de salida de un cuantizador uniforme, para un tono de prueba a plena carga, es

$$(SNR)_O = \frac{A_m^2/2}{A_m^2 2^{-2R}/3} = \frac{3}{2}(2^{2R}) \tag{3.34}$$

Al expresar la relación señal a ruido en decibeles, tenemos

$$10 \log_{10}(\text{SNR})_O = 1.8 + 6R \tag{3.35}$$

TABLA 3.1 ***Relación señal a ruido (cuantización) para un número variable de niveles de representación correspondiente a modulación senoidal***

Número de niveles de representación, L	*Número de bits por muestra*, R	*Relación señal a ruido (dB)*
32	5	31.8
64	6	37.8
128	7	43.8
256	8	49.8

Para diversos valores de L y R, los valores correspondientes de la relación señal a ruido se presentan en la tabla 3.1. A partir de la tabla es posible realizar una rápida estimación del número de bits por muestra que se requiere para una relación señal a ruido de salida deseada, suponiendo modulación senoidal. ◀

Hasta ahora en esta sección nos hemos centrado en cómo caracterizar cuantizadores escalares sin memoria y en evaluar su desempeño. Sin embargo, al hacerlo de ese modo evitamos el diseño óptimo de los cuantizadores, esto es, el aspecto de elegir los niveles de representación y las celdas de partición de manera que se minimice la potencia promedio de cuantización para un número preestablecido de niveles de representación. Desafortunadamente, este problema de optimización no conduce por sí solo a una solución de forma cerrada, en virtud de la naturaleza altamente *no lineal* del proceso de cuantización. Más bien, contamos con algoritmos efectivos para determinar el diseño óptimo de manera iterativa. Un algoritmo bien conocido que merece mencionarse en este contexto es el cuantizador de Lloyd-Max, que se discute a continuación.

■ CONDICIONES PARA LA OPTIMIZACIÓN DE CANTIDADES ESCALARES

El desafío que se enfrenta al diseñar un cuantizador escalar es cómo elegir los niveles de representación y las celdas de partición en torno a ellos, de modo que se minimice la potencia promedio de cuantización para un número fijo de niveles de representación.

Para establecer el problema en términos matemáticos, considere una señal de mensaje $m(t)$ extraída de un proceso estacionario $M(t)$. Dejemos que $-A \leq m \leq A$ denote el intervalo dinámico de $m(t)$, el cual se divide en un conjunto de L celdas, como se indica en la figura 3.12. Las fronteras de las celdas de partición se definen mediante un conjunto de números reales $m_1, m_2, \ldots, m_{L+1}$ que satisfacen las siguientes tres condiciones:

$$\begin{aligned} m_1 &= -A \\ m_{L+1} &= A \\ m_k &\leq m_{k+1} \text{ para } k = 1, 2, \ldots, L \end{aligned}$$

La celda de partición k-ésima está definida por

$$\mathcal{J}_k: m_k < m \leq m_{k+1} \text{ para } k = 1, 2, \ldots, L \tag{3.36}$$

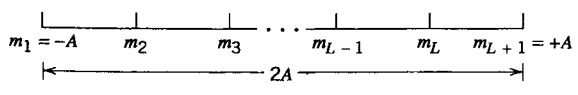


FIGURA 3.12 Ilustración de la división del intervalo dinámico $-A \leq m \leq A$ de una señal de mensaje $m(t)$ en un conjunto de L celdas.

Denotemos los niveles de representación (es decir, los valores de cuantización) por v_k, $k = 1, 2, \ldots, L$. Entonces, suponiendo que $d(m, v_k)$ denota una *medida de la distorsión* al usar v_k para representar todos aquellos valores de la entrada m que se encuentran dentro de la celda de partición $\mathcal{J}_k$, la meta es determinar dos conjuntos, $\{v_k\}_{k=1}^{L}$ y $\{\mathcal{J}_k\}_{k=1}^{L}$, que minimicen la *distorsión promedio*

$$D = \sum_{k=1}^{L} \int_{m \in \mathcal{J}_k} d(m, v_k) f_M(m)\, dm \tag{3.37}$$

donde $f_M(m)$ es la función de densidad de probabilidad de la variable aleatoria M con valor de muestreo m.

Una medida de distorsión que se usa comúnmente es

$$d(m, v_k) = (m - v_k)^2 \tag{3.38}$$

en cuyo caso hablamos de la *distorsión media cuadrática*. En cualquier caso, el problema de optimización planteado aquí es no lineal, definiéndose una solución explícita y de forma cerrada. Para sortear esta dificultad, recurrimos a un método algorítmico para resolver el problema de manera iterativa.

En términos estructurales, el cuantizador consiste en dos componentes con parámetros de diseño interrelacionados:

- Un codificador caracterizado por un conjunto de celdas de partición $\{\mathcal{J}_k\}_{k=1}^{L}$; éste se ubica en el transmisor.
- Un decodificador caracterizado por un conjunto de niveles de representación $\{v_k\}_{k=1}^{L}$; éste se localiza en el receptor.

En consecuencia, es posible identificar dos condiciones críticamente importantes que proporcionan las bases matemáticas para todas las soluciones algorítmicas al problema de la cuantización óptima. Una condición supone que estamos dando un decodificador y que el problema radica en determinar el codificador óptimo en el transmisor. La otra condición considera que estamos dando un codificador y que el problema reside en determinar el decodificador óptimo en el receptor. De aquí en adelante, estas dos condiciones se indicarán, respectivamente, como la condición I y la condición II.

Condición I. Optimización del codificador para un decodificador dado

La disponibilidad de un decodificador quiere decir que tenemos cierto *libro de códigos* en mente. Considere que el libro de códigos se define mediante

$$\mathcal{C} : \{v_k\}_{k=1}^{L} \tag{3.39}$$

Dado el libro de códigos $\mathcal{C}$, el problema es determinar el conjunto de celdas de partición $\{\mathcal{J}_k\}_{k=1}^{L}$ que minimice la distorsión promedio D. Es decir, deseamos encontrar el codificador definido por la correspondencia no lineal

$$g(m) = v_k, \qquad k = 1, 2, \ldots, L \tag{3.40}$$

tal que tengamos

$$D = \int_{-A}^{A} d(m, g(m)) f_M(m)\, dM \geq \sum_{k=1}^{L} \int_{m \in \mathcal{J}_k} [\min_{v_k \in \mathcal{C}} d(m, v_k)]\, f_M(m)\, dm \tag{3.41}$$

Para que se alcance la cota inferior que se especifica en la ecuación (3.41) requerimos que la correspondencia no lineal de la ecuación (3.40) se cumpla sólo si la condición

$$d(m, v_k) \leq d(m, v_j) \quad \text{se cumple para toda } j \neq k \tag{3.42}$$

La condición necesaria que se describe en la ecuación (3.42) para la optimización del codificador de un libro de códigos especificado $\mathscr{C}$ se reconoce como la *condición del vecino más cercano.* En otras palabras, la condición del vecino más cercano requiere que la celda de partición $\mathscr{J}_k$ incorpore todos aquellos valores de la entrada m que están más cercanos a v_k que cualquier otro elemento del libro de códigos $\mathscr{C}$. Esta condición de optimización se satisface en realidad en forma intuitiva.

Condición II. Optimización del decodificador para un codificador dado

Considere la siguiente situación inversa a la que se describió de acuerdo con la condición I, la cual puede establecerse del siguiente modo: optimice el libro de códigos $\mathscr{C} = \{v_k\}_{k=1}^{L}$ para el decodificador, dado que el conjunto de celdas de partición $\{\mathscr{J}_k\}_{k=1}^{L}$ que caracteriza al codificador es fijo. El criterio para la optimización es la distorsión promedio (cuadrática media):

$$D = \sum_{k=1}^{L} \int_{m \in \mathscr{J}_k} (m - v_k)^2 f_M(m)\, dm \tag{3.43}$$

La función de densidad de probabilidad $f_M(m)$ es claramente independiente del libro de códigos $\mathscr{C}$. Por tanto, al diferenciar D con respecto al nivel de representación v_k, obtenemos de inmediato

$$\frac{\partial D}{\partial v_k} = -2 \sum_{k=1}^{L} \int_{m \in \mathscr{J}_k} (m - v_k) f_M(m)\, dm \tag{3.44}$$

Al igualar a cero $\partial D/\partial v_k$ y resolver para v_k obtenemos el valor óptimo

$$v_{k,\text{opt}} = \frac{\int_{m \in \mathscr{J}_k} m f_M(m)\, dm}{\int_{m \in \mathscr{J}_k} f_M(m)\, dm}$$

El denominador de la ecuación (3.45) es justamente la probabilidad, p_k, de que la variable aleatoria m con valor de muestreo M se encuentre en la celda de partición $\mathscr{J}_k$, como se indica por medio de

$$\begin{aligned} p_k &= P(m_k < M \leq m_k + 1) \\ &= \int_{m \in \mathscr{J}_k} f_M(m)\, dm \end{aligned} \tag{3.46}$$

Por tanto, es posible interpretar la condición de optimización de la ecuación (3.45) como la elección del nivel de representación v_k igual a la *media condicional* de la variable aleatoria M, dado que M se encuentra en la celda de partición $\mathscr{J}_k$. De este modo podemos establecer formalmente la condición de optimización del decodificador para un codificador dado en la forma siguiente:

$$v_{k,\text{ opt}} = E[M \mid m_k < M \leq m_{k+1}] \tag{3.47}$$

donde E es el operador de esperanza. La ecuación (3.47) también se satisface en forma intuitiva.

Advierta que la condición del vecino más cercano (condición I) para la optimalidad del codificador para un decodificador determinado se probó para una distorsión promedio genérica. Sin embargo, el requerimiento de la media condicional (condición II) para la optimalidad del decodificador relativa a un codificador determinado se probó para el caso especial de una distorsión cuadrática media. En cualquier caso, estas dos condiciones son necesarias para la optimalidad de un cuantizador escalar. Básicamente, el algoritmo para diseñar el cuantizador consiste en la optimización alternativa del codificador de acuerdo con la condición I, en la optimización posterior del decodificador en apego a la condición II, y en proseguir de este modo

hasta que la distorsión promedio D alcance un mínimo. Un cuantizador óptimo diseñado de este modo recibe el nombre de *cuantizador de Lloyd-Max.*[4]

3.7 *Modulación por codificación de pulsos*

Con los procesos de muestreo y cuantización a nuestra disposición, estamos listos para describir la modulación por codificación de pulsos, la cual, como se mencionó antes, constituye la forma más fundamental de la modulación digital por pulsos. En la *modulación por codificación de pulsos (PCM), una señal de mensaje se representa por medio de una secuencia de pulsos codificados, la cual se consigue al representar la señal en forma discreta tanto en tiempo como en amplitud.* Las operaciones básicas que se efectúan en el transmisor de un sistema PCM son el *muestreo, la cuantización* y *la codificación,* como se muestra en la figura 3.13*a*; el filtro pasobajas anterior al muestreo se incluye para evitar el traslape de la señal de mensaje. Las operaciones de cuantización y codificación se ejecutan casi siempre en el mismo circuito, el cual recibe el nombre de *convertidor analógico-digital.* Las operaciones básicas en el receptor son la *regeneración* de señales deterioradas, la *decodificación* y la *reconstrucción* del tren de muestras cuantizadas, como se ilustra en la figura 3.13*c*. La regeneración ocurre también en puntos intermedios a lo largo de la trayectoria de transmisión según sea necesario, como se señala en la figura 3.13*b*. Cuando se recurre al multiplexado por división del tiempo, se vuelve necesario sincronizar el receptor con el transmisor para que todo el sistema opere en forma satisfactoria, como se explica en la sección 3.9. En lo que sigue describimos las diversas operaciones que constituyen un sistema PCM básico.

Muestreo

La señal de mensaje entrante se muestrea con un tren de pulsos rectangulares angostos de manera que se aproxime bastante al proceso de muestreo instantáneo. Para asegurar la perfecta reconstrucción de la señal de mensaje en el receptor, la frecuencia de muestreo debe ser mayor que el doble de la componente de frecuencia más alta W de la señal de mensaje de acuerdo con el teorema de muestreo. En la práctica se utiliza un filtro antitraslape pasobajas en el extremo frontal del muestreador para excluir las frecuencias superiores a W antes del muestreo. De tal

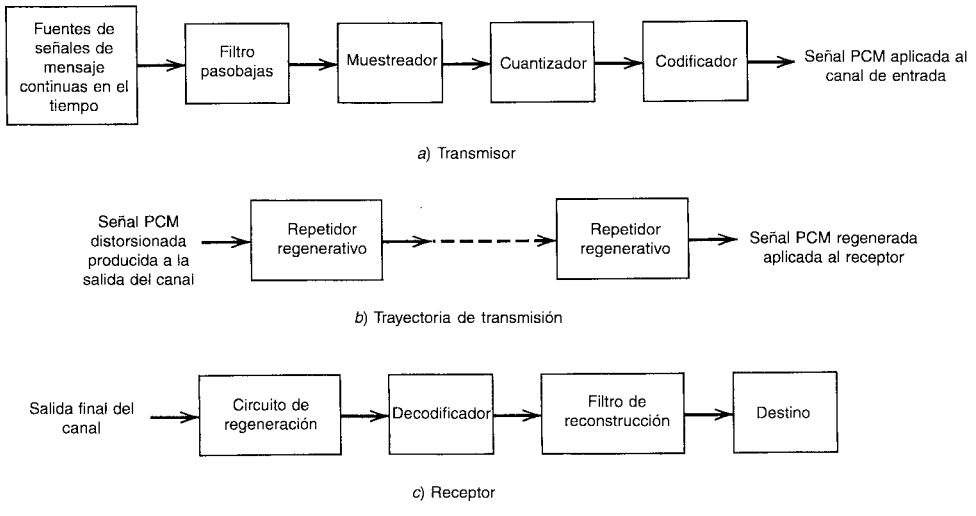


FIGURA 3.13 Los elementos básicos de un sistema PCM.

modo, la aplicación del muestreo permite la reducción de la señal de mensaje que varía continuamente (de alguna duración finita) a un número limitado de valores discretos por segundo.

■ CUANTIZACIÓN

La versión muestreada de la señal de mensaje se cuantiza después proporcionando de ese modo una nueva representación de la señal que es discreta tanto en tiempo como en amplitud. El proceso de cuantización puede seguir una ley uniforme, como se describe en la sección 3.6. Sin embargo, en las comunicaciones telefónicas resulta preferible utilizar una separación de variables entre los niveles de representación. Por ejemplo, el intervalo de voltajes que abarcan las señales de voz, desde los picos del habla fuerte hasta los pasajes débiles del habla poco intensa, es del orden de 1000 a 1. Al utilizar un *cuantizador no uniforme* con la característica de que el tamaño de escalón aumenta cuando se incrementa la separación a partir del origen de la característica de amplitud de entrada-salida, los grandes escalones finales del cuantizador pueden ocuparse de posibles excursiones de la señal de voz hacia grandes intervalos de amplitud que ocurren relativamente poco. En otras palabras, los pasajes débiles que necesitan más protección, se favorecen a costa de los pasajes ruidosos. De este modo se consigue una precisión porcentual casi uniforme a través de la mayor parte del intervalo de amplitud de la señal de entrada, resultando que se necesitan muy pocos escalones, lo cual sería al contrario si se utilizara un cuantizador uniforme.

El uso de un cuantizador no uniforme es equivalente a hacer pasar la señal de la banda base por un *compresor* y aplicar después la señal comprimida a un cuantizador uniforme. Una forma particular de ley de compresión que se usa en la práctica es la llamada *ley* μ,[5] la cual se define por medio de

$$|v| = \frac{\log(1+\mu|m|)}{\log(1+\mu)} \tag{3.48}$$

donde m y v son los voltajes de entrada y salida normalizados, y μ es una constante positiva. En la figura 3.14*a* hemos graficado la ley μ para tres valores diferentes de μ. El caso de cuantización uniforme corresponde a $\mu = 0$. Para un valor de μ determinado, la pendiente recíproca de la

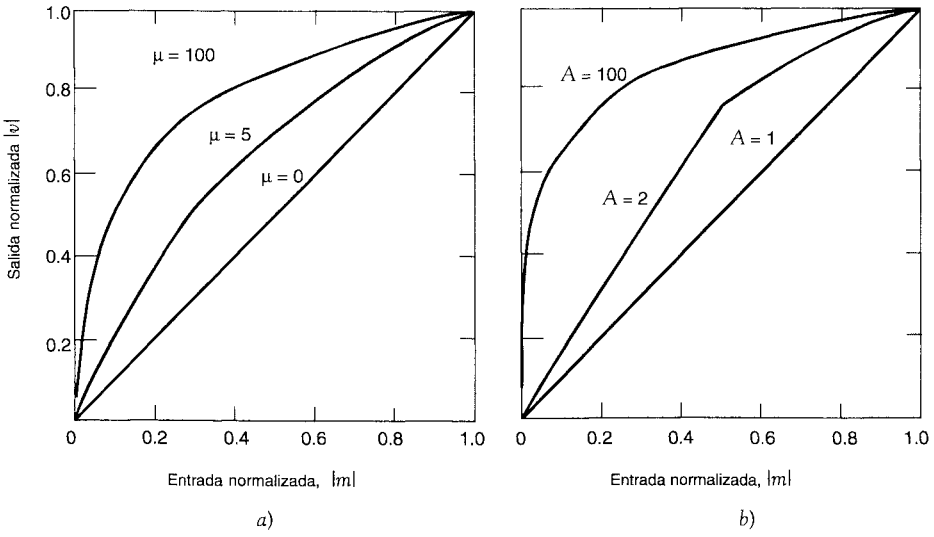


FIGURA 3.14 Leyes de compresión. *a*) Ley μ. *b*) Ley A.

curva de compresión que define los escalones cuánticos, está dada por la derivada de $|m|$ con respecto a $|v|$, esto es,

$$\frac{d|m|}{d|v|} = \frac{\log(1+\mu)}{\mu}(1+\mu|m|) \tag{3.49}$$

Por tanto, vemos que la ley μ no es ni estrictamente lineal ni estrictamente logarítmica, sino que es lineal de modo aproximado a niveles de entrada bajos correspondientes a $\mu|m| << 1$, y casi logarítmica a niveles de entrada elevados correspondientes a $\mu|m| >> 1$.

Otra ley de compresión que se usa en la práctica es la denominada *ley A* definida por

$$|v| = \begin{cases} \dfrac{A|m|}{1+\log A}, & 0 \le |m| \le \dfrac{1}{A} \\ \dfrac{1+\log(A|m|)}{1+\log A}, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{3.50}$$

que se grafica en la figura 3.14*b* para A variable. El caso de la cuantización uniforme corresponde a $A = 1$. La pendiente recíproca de esta segunda curva de compresión está dada por la derivada de $|m|$ con respecto a $|v|$, como indica (dependiendo del valor que se asigna a la entrada normalizada $|m|$)

$$\frac{d|m|}{d|v|} = \begin{cases} \dfrac{1+\log A}{A}, & 0 \le |m| \le \dfrac{1}{A} \\ (1+A)|m|, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \tag{3.51}$$

Para regresar las muestras de señal a su nivel relativo correcto, debemos utilizar un dispositivo en el receptor con una característica complementaria a la del compresor. Dicho dispositivo recibe el nombre de *expansor*. Idealmente, las leyes de compresión y expansión son en sentido opuesto, por lo que, salvo por el efecto de cuantización, la salida del expansor es igual a la entrada del compresor. La combinación de un *comp*resor y un exp*ansor* recibe el nombre de *compandor*.

Tanto para la ley μ como para la ley A, la capacidad del intervalo dinámico del compandor mejora con el incremento de μ y A, respectivamente. La SNR para señales de bajo nivel aumenta a costa de la SNR correspondiente a las señales de alto nivel. Para ajustar estos dos requerimientos en conflicto (es decir, una relación señal a ruido razonable tanto para señales de bajo como de alto nivel), suele establecerse un compromiso al elegir el valor del parámetro μ para la ley μ y el parámetro A para la ley A. Los valores representativos que se utilizan en la práctica son $\mu = 255$ y $A = 87.6$.

También es de interés advertir que en los sistemas PCM reales, la circuitería de compresión y expansión no produce una réplica exacta de las curvas de compresión no lineales que se muestran en la figura 3.14. En vez de eso, proporcionan una aproximación *lineal por secciones* a la curva deseada. Utilizando un número suficientemente grande de segmentos lineales, la aproximación puede acercarse en un alto grado a la curva de compresión real. Esta forma de aproximación se ilustra en el ejemplo 3.2.

Codificación

Al combinar los procesos de muestreo y cuantización, la especificación de una señal de mensaje (banda base) se limita a un conjunto discreto de valores, aunque no es la forma más apropiada para la transmisión por una línea telefónica o una trayectoria de radio. Para explotar las ventajas del muestreo y la cuantización con el fin de lograr que la señal transmitida sea más robusta con respecto al ruido, la interferencia y otros deterioros del canal, requerimos el uso de un

TABLA 3.2 *Sistema numérico binario para* **R = 4 bits/muestra**

Número ordinal del nivel de representación	*Número del nivel expresado como suma de potencias de 2*	*Número binario*
0		0000
1	2^0	0001
2	2^1	0010
3	$2^1 + 2^0$	0011
4	2^2	0100
5	$2^2 + 2^0$	0101
6	$2^2 + 2^1$	0110
7	$2^2 + 2^1 + 2^0$	0111
8	2^3	1000
9	$2^3 + 2^0$	1001
10	$2^3 + 2^1$	1010
11	$2^3 + 2^1 + 2^0$	1011
12	$2^3 + 2^2$	1100
13	$2^3 + 2^2 + 2^0$	1101
14	$2^3 + 2^2 + 2^1$	1110
15	$2^3 + 2^2 + 2^1 + 2^0$	1111

proceso de codificación para trasladar el conjunto discreto de valores de muestreo a una forma más apropiada de señal. Cualquier plan para representar cada uno de este conjunto de valores discretos como un arreglo particular de eventos discretos recibe el nombre de *código*. Uno de los eventos discretos en un código se denomina *elemento de código* o *símbolo*. Por ejemplo, la presencia o ausencia de un pulso es un símbolo. Un arreglo particular de símbolos utilizado en un código para representar un valor único del conjunto discreto recibe el nombre de *palabra de código* o *carácter*.

En un *código binario*, cada símbolo puede tener dos valores o tipos distintos, como la presencia o ausencia de un pulso. Los dos símbolos de un código binario suelen denotarse como 0 y 1. En un *código ternario*, cada símbolo puede ser uno de tres valores o tipos de valores distintos, y así sucesivamente para otros códigos. Sin embargo, *la máxima ventaja sobre los efectos de ruido en un medio de transmisión se obtiene utilizando un código binario, debido a que un símbolo binario soporta un nivel relativamente alto de ruido y es fácil de regenerar.* Suponga que, en un código binario, cada palabra del código consiste en R bits: *bit* es el acrónimo de las palabras en inglés *binary digit* (dígito binario); de tal modo, R denota el número de *bits por muestra*. Entonces, utilizando un código de este tipo, es posible representar un total de 2^R números distintos. Por ejemplo, una muestra cuantizada en uno de 256 niveles puede representarse mediante una palabra de código de 8 bits.

Hay varias maneras de establecer una correspondencia uno a uno entre niveles de representación y palabras de código. Un método conveniente es expresar el número ordinal del nivel de representación como un número binario. En el sistema de números binarios, cada dígito tiene un valor de lugar que es una potencia de 2, como se ilustra en la tabla 3.2 para el caso de 4 bits por muestra (es decir, $R = 4$).

Códigos de línea

Es posible utilizar cualesquiera de varios códigos de línea para la representación eléctrica de una secuencia binaria de datos. La figura 3.15 presenta las formas de onda de cinco importantes códigos de línea para la secuencia de datos de ejemplo 01101001. La figura 3.16 ilustra

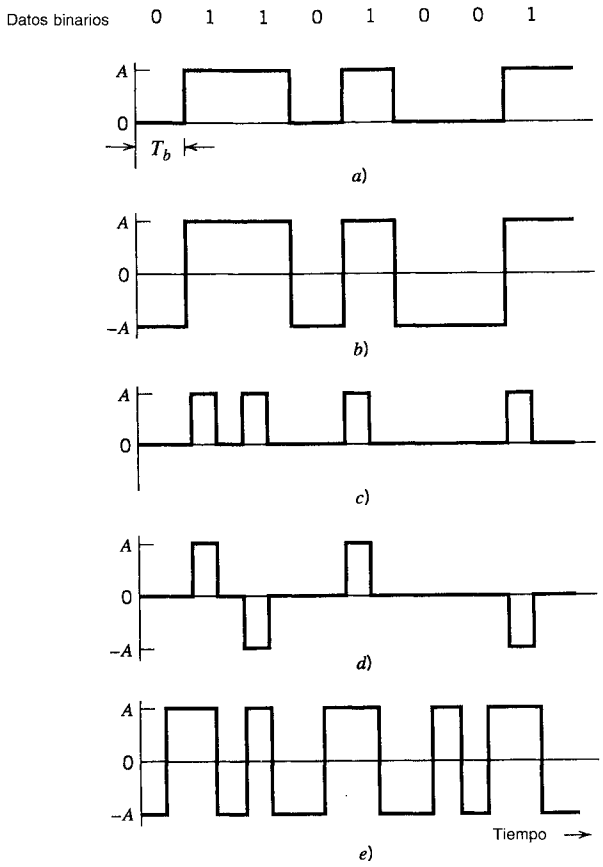


FIGURA 3.15 Códigos de línea para la representación eléctrica de datos binarios. *a*) Transmisión de señales unipolares sin retorno a cero. *b*) Transmisión de señales polares sin retorno a cero. *c*) Señales unipolares con retorno a cero. *d*) Transmisión bipolar con retorno por cero. *e*) Código de división de fase o de Manchester.

sus espectros de potencia individuales (para frecuencias positivas) correspondientes a datos binarios generados al azar, suponiendo que 1) los símbolos 0 y 1 son igualmente probables, 2) la potencia en promedio está normalizada a la unidad y 3) la frecuencia f se normaliza con respecto a la tasa de bits $1/T_b$. (En cuanto a las fórmulas que se utilizaron para graficar los espectros de potencia de la figura 3.16, se remite al lector al problema 3.11.) Los cinco códigos de línea que se ilustran en la figura 3.15 se describen a continuación:

1. *Transmisión de señales unipolares sin retorno a cero (NRZ)*
En este código, el símbolo 1 se representa transmitiendo un pulso de amplitud A para la duración del símbolo y el símbolo 0 se representa al desactivar el pulso como en la figura 3.15*a*. Este código de línea se conoce también como *transmisión de señales activación-desactivación*. Las desventajas de ésta son el desperdicio de potencia debido al nivel de cd transmitido y al hecho de que el espectro de potencia de la señal que se transmite no se acerca a cero a la frecuencia cero.

2. *Transmisión de señales polares sin retorno a cero (NRZ)*
En este segundo código de línea, los símbolos con número 1 y 0 se representan al transmitir pulsos de amplitudes $+A$ y $-A$, respectivamente, como se indica en la figura 3.15*b*. Este código de línea es fácil de generar, aunque su desventaja es que el espectro de la señal es más grande cerca de la potencia cero.

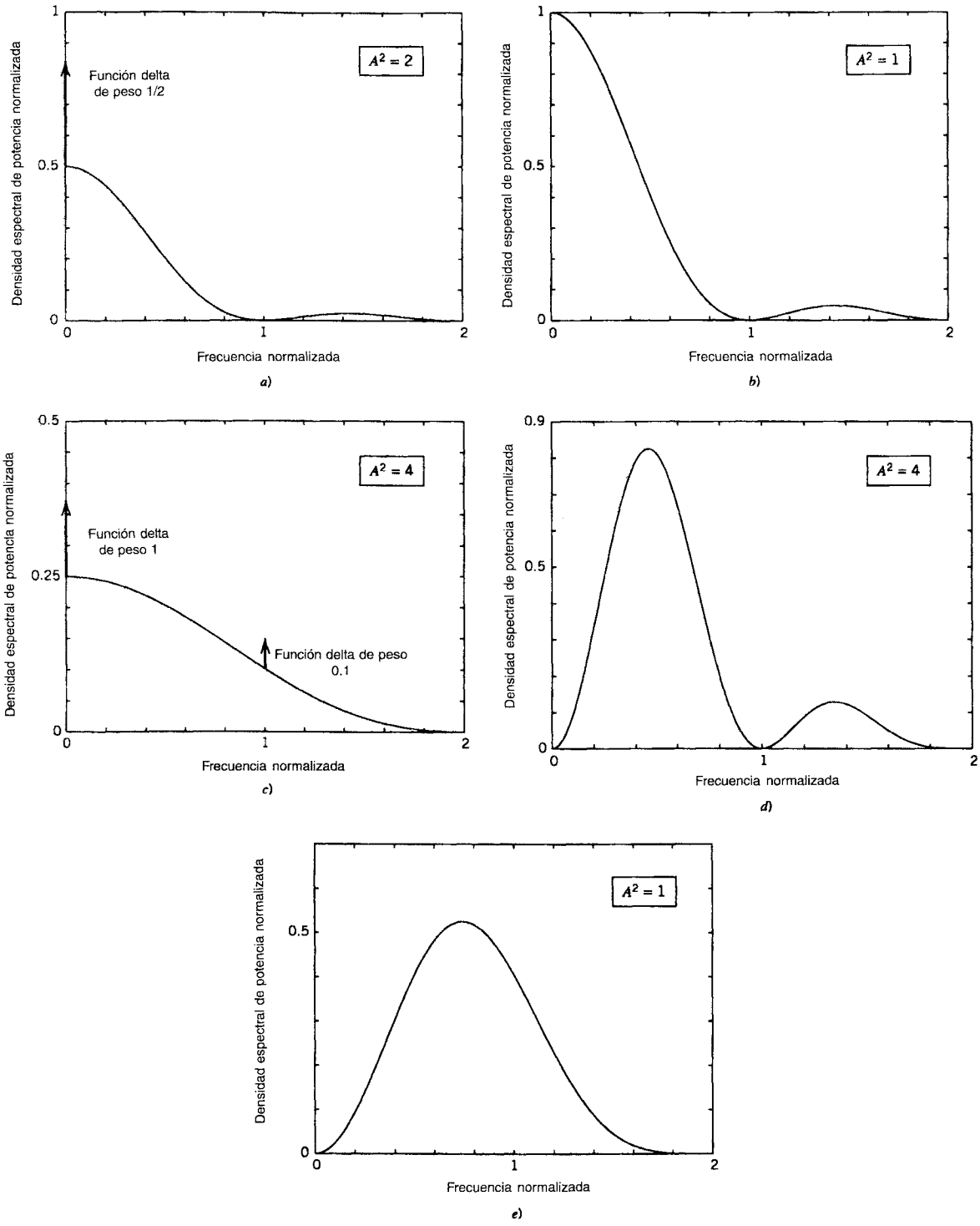


FIGURA 3.16 Espectro de potencia de códigos de línea: *a*) Señal unipolar sin retorno a cero. *b*) Señal polar sin retorno a cero. *c*) Señal unipolar con retorno a cero. *d*) Señal bipolar con retorno a cero. *e*) Señal con codificación de Manchester. La frecuencia se normaliza con respecto a la tasa de bits $1/T_b$, y la potencia promedio se normaliza en la unidad.

3. *Transmisión de señales unipolares con retorno a cero (RZ)*
En este otro código de línea, el símbolo 1 se representa mediante un pulso rectangular de amplitud A y ancho de medio símbolo, y el símbolo 0 se representa al *no* transmitir pulso, como se ilustra en la figura 3.15c. Un rasgo atractivo de este código de línea es la presencia de las funciones delta en $f = 0, \pm 1/T_b$ en el espectro de potencia de la señal transmitida, y que puede utilizarse para recuperar la temporización de bits en el receptor. Sin embargo, su desventaja es que requiere 3 dB más de potencia que la transmisión polar con retorno a cero para la misma probabilidad del error de símbolo; este aspecto se considera en el capítulo 4 como parte del problema 4.10.

4. *Transmisión de señales bipolares con retorno a cero (BRZ)*
Este código de línea recurre a tres niveles de amplitud, como se indica en la figura 3.15*d*. De modo específico, se utilizan pulsos positivo y negativo de igual amplitud (es decir, +A y − A) de modo alternativo con el símbolo 1, teniendo cada pulso la mitad del ancho del símbolo; nunca se usa un pulso para el símbolo 0. Una propiedad útil de la transmisión de señales BRZ es que el espectro de potencia de la señal transmitida no tiene componente de cd y componentes de baja frecuencia relativamente insignificantes para el caso cuando ocurren con igual probabilidad los símbolos 1 y 0. Esta línea de código se conoce también como *transmisión de señales de inversión de marca alternada* (AMI).

5. *Fase dividida (código Manchester)*
En este método de transmisión de señales, que se ilustra en la figura 3.15*e*, el símbolo 1 se representa mediante un pulso positivo de amplitud A seguido por un pulso negativo de amplitud − A, con ambos pulsos con ancho de la mitad del símbolo. Para el símbolo 0 se invierten las polaridades de estos dos pulsos. El código Manchester suprime la componente de cd y tiene componentes de baja frecuencia relativamente insignificantes, a pesar de las estadísticas de la señal. Esta propiedad resulta esencial en algunas aplicaciones.

Codificación diferencial

Este método se emplea para codificar información en términos de *transiciones de señal*. En particular, se utiliza una transición para designar el símbolo 0 en la secuencia de datos binarios entrante, en tanto que la no transmisión se utiliza para designar el símbolo 1, como se ilustra en la figura 3.17. En la figura 3.17*b* mostramos la secuencia de datos codificados diferencialmente para los datos del ejemplo especificado en la figura 3.17*a*. La secuencia original de datos binarios que se emplea en este caso es la misma que se utilizó en la figura 3.15. La forma de onda de los datos codificados diferencialmente se muestra en la figura 3.17*c*, suponiendo el uso de transmisión de señales unipolares sin retorno a cero. A partir de la figura 3.17 es claro que una señal codificada en forma diferencial puede invertirse sin afectar su interpretación. La información binaria original se recupera comparando simplemente la polaridad de los símbolos binarios adyacentes para establecer si ha ocurrido o no una transición. Advierta que la codificación diferencial requiere el uso de un *bit de referencia* antes de iniciar el proceso de codificación. En la figura 3.17 se utiliza el símbolo 1 como bit de referencia.

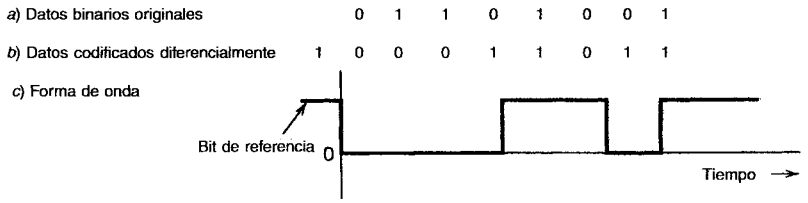


FIGURA 3.17 *a*) Datos binarios originales. *b*) Datos codificados diferencialmente, suponiendo el bit de referencia 1. *c*) Forma de onda de los datos codificados diferencialmente utilizando transmisión de señales unipolares sin retorno a cero.

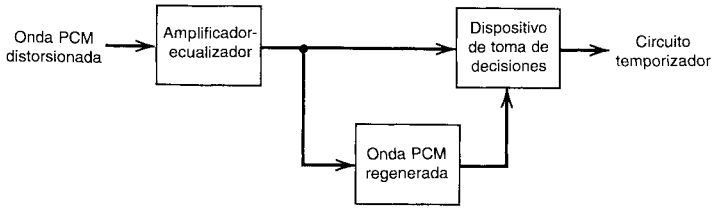


FIGURA 3.18 Diagrama de bloques del repetidor regenerativo.

■ REGENERACIÓN

El rasgo más importante de los sistemas PCM se encuentra en la capacidad para controlar los efectos de distorsión y ruido que se producen al transmitir una señal PCM por un canal. Esta capacidad se logra al construir la señal PCM mediante una cadena de *repetidores regenerativos* ubicados a una distancia suficientemente próxima a lo largo de la ruta de transmisión. Como se ilustra en la figura 3.18, un repetidor regenerativo lleva a cabo tres funciones básicas: *ecualización, temporización* y *toma de decisiones.* El ecualizador da forma a los pulsos recibidos de tal manera que compensa los efectos de las distorsiones de amplitud y fase que producen las características de transmisión no ideales del canal. El circuito de temporización proporciona un tren de pulsos periódicos, obtenido de los pulsos recibidos, para muestrear los pulsos ecualizados en los instantes en los que la relación señal a ruido es un máximo. Cada muestra extraída de esa manera se compara con un *umbral* predeterminado en el dispositivo de toma de decisiones. En cada intervalo de bit se efectúa entonces una decisión si el símbolo que se recibe es un 1 o un 0, según se supere o no el umbral. Si se supera el umbral, un nuevo pulso de limpieza que representa al símbolo 1 se transmite al siguiente repetidor. En otro caso, se transmite otro pulso nuevo de limpieza que representa al símbolo 0 al siguiente repetidor. De esta manera se elimina por completo la acumulación de la distorsión y el ruido en un tramo de repetidor, siempre y cuando la perturbación no sea demasiado grande para provocar un error en el proceso de toma de decisiones. De modo ideal, salvo por un retardo, la señal regenerada es la misma que se transmite originalmente. Sin embargo, en la práctica la señal regenerada se aparta de la original por dos razones principales:

1. La inevitable presencia del ruido y la interferencia del canal ocasionan que el repetidor tome ocasionalmente decisiones equivocadas, introduciendo *errores de bit* en la señal regenerada.
2. Si el espaciamiento entre los pulsos recibidos se desvía de su valor asignado, se introduce una *inestabilidad* en la posición del pulso regenerado, provocando por ello distorsión.

■ DECODIFICACIÓN

La primera operación en el receptor es regenerar (es decir, reformar y limpiar) los pulsos recibidos anteriormente. Estos pulsos limpios se reagrupan en palabras de código y se decodifican (es decir, se corresponden en sentido inverso) en una señal PAM cuantizada. El proceso de *decodificación* implica generar un pulso cuya amplitud es la suma lineal de todos los pulsos en la palabra de código, siendo cada pulso ponderado por su lugar (2^0, 2^1, 2^2, ..., 2^{R-1}) en el código, donde R es el número de bits por muestra.

■ FILTRADO

La operación final en el receptor consiste en recuperar la señal de mensaje al pasar la salida del decodificador por un filtro de reconstrucción pasobajas cuya frecuencia de corte es igual al ancho de

banda W del mensaje. Suponiendo que la trayectoria de transmisión es sin errores, la señal recuperada no incluye ruido, con la excepción de la distorsión inicial que introduce el proceso de cuantización.

3.8 Consideraciones de ruido en sistemas PCM

El desempeño de un sistema PCM resulta afectado por dos fuentes principales de ruido:

1. *Ruido del canal,* que se presenta en cualquier lugar entre la salida del transmisor y la entrada del receptor. El ruido del canal siempre está presente luego de que se activa el equipo.
2. *Ruido de cuantización,* el cual se presenta en el transmisor y se transporta en todo momento hasta la salida del receptor. A diferencia del ruido del canal, el ruido de cuantización *depende de la señal,* pues desaparece cuando se corta la señal del mensaje.

Naturalmente, estas dos fuentes de ruido aparecen en forma simultánea una vez que el sistema PCM está en operación. Sin embargo, la práctica tradicional es considerarlas por separado, de manera que se pueda obtener información amplia acerca de sus efectos individuales en el desempeño del sistema.

El principal efecto del canal de ruido radica en introducir *errores de bit* en la señal recibida. En el caso de un sistema PCM binario, la presencia de un error de bit provoca que el símbolo 1 se confunda con el símbolo 0 o viceversa. Sin duda, cuanto más frecuentemente se presenten los errores de bit, tanto más distinta se vuelve la salida del receptor en comparación con la señal del mensaje original. La fidelidad de la transmisión de información mediante PCM en la presencia de ruido de canal puede medirse en términos de la *probabilidad promedio del error de símbolo,* la cual se define como la probabilidad de que el símbolo reconstruido en la salida del receptor difiera del símbolo binario transmitido, en promedio. La probabilidad promedio del error de símbolo, conocida también como *tasa de error de bits* (BER), supone que todos los bits en la onda binaria original son de igual importancia. Sin embargo, cuando existe mayor interés en reconstruir la forma de onda analógica de la señal de mensaje original, los errores de símbolo diferentes quizá necesiten *ponderarse* de manera distinta; por ejemplo, un error en el bit más significativo en una palabra de código (que representa a una muestra cuantizada de una señal de mensaje) es más peligroso que un error en el bit menos significativo.

Para optimizar el desempeño del sistema en la presencia de ruido del canal, necesitamos minimizar la probabilidad promedio del error de símbolo. Para esta evaluación, resulta usual modelar el ruido del canal como gaussiano, aditivo y blanco. El efecto del ruido del canal puede hacerse prácticamente ignorable al asegurar el uso de una adecuada relación de energía de la señal a densidad de ruido mediante la provisión de espaciamiento suficientemente corto entre los repetidores regenerativos en el sistema PCM. En una situación de este tipo, el desempeño del sistema PCM está limitado esencial y exclusivamente por el ruido de cuantización.

A partir de la explicación del ruido de cuantización que se presentó en la sección 3.6, admitimos que este tipo de ruido se encuentra en esencia bajo el control del diseñador. Puede hacerse tan pequeño como para ignorarse mediante el uso adecuado de niveles de representación en el cuantizador y de la selección de una estrategia de compresión y expansión acoplada a las características del tipo de señal de mensaje que se transmite. De ese modo, encontramos que el uso de la PCM ofrece la posibilidad de construir un sistema que sea *resistente* al ruido del canal en una escala mayor que la capacidad de cualquier modulación de onda continua o sistema de modulación analógica por pulsos.

Umbral de error

La teoría que fundamenta el cálculo de la tasa de error de bits en un sistema PCM se posterga hasta el capítulo 4. Por ahora, es suficiente afirmar que la probabilidad promedio del error de símbolo en un receptor PCM de codificación binaria debida a ruido aditivo blanco gaussiano

depende exclusivamente de E_b/N_0, lo cual se define como *la proporción entre la energía de la señal transmitida por, bit* E_b, *y la densidad espectral de ruido,* N_0. Advierta que la proporción E_b/N_0 es adimensional aun cuando las cantidades E_b y N_0 tienen diferente significado físico. En la tabla 3.3 presentamos un resumen de esta dependencia para el caso de un sistema PCM binario que utiliza transmisión de señales polar sin retorno a cero. El resultado que se presenta en la última columna de la tabla supone una tasa de bits de 10^5 b/s.

De acuerdo con la tabla 3.3 es claro que existe un *umbral de error* (a casi 11 dB). Para E_b/N_0 debajo del umbral de error el desempeño del receptor implica números importantes de errores, y sobre el umbral el efecto del ruido del canal es prácticamente imperceptible. En otras palabras, siempre que el cociente E_b/N_0 supere el umbral de error, el ruido del canal no tiene virtualmente efectos sobre el desempeño del receptor, lo cual es precisamente la meta de la PCM. Sin embargo, cuando E_b/N_0 disminuye por debajo del umbral de error, existe un aumento agudo en la velocidad a la cual ocurren los errores en el receptor. Debido a que los errores de decisión originan la construcción de palabras de código incorrectas, encontramos que cuando los errores son frecuentes, el mensaje que se reconstruye en la salida del receptor contiene poca semejanza con el mensaje original.

Al comparar la figura de 11 dB para el umbral de error en un sistema PCM utilizando transmisión de señales polares NRZ con los 60-70 dB que se requieren para la transmisión de habla de alta calidad utilizando modulación de amplitud, vemos que la PCM requiere mucho menos potencia, aun cuando la potencia promedio del ruido en el sistema PCM se aumente R veces en ancho de banda, donde R es el número de bits en una palabra de código (es decir, bits por muestra).

En la mayoría de los sistemas de transmisión se acumulan los efectos del ruido y la distorsión que provienen de los enlaces individuales. Para una calidad determinada de la transmisión completa, cuanto más larga resulta la separación física entre el transmisor y el receptor, tanto más estrictos se vuelven los requerimientos en cada enlace del sistema. Sin embargo, en un sistema PCM, puesto que la señal puede regenerarse tan a menudo como sea necesario, los efectos de la amplitud, la fase y las distorsiones en un enlace (si no son demasiado severos) prácticamente no tienen efecto en la señal regenerada para el siguiente enlace. También observamos que el efecto del ruido del canal puede hacerse casi ignorable utilizando una proporción E_b/N_0 arriba del umbral. Para todos los fines prácticos, los requerimientos de transmisión para un enlace PCM son casi independientes de la longitud física del canal de comunicación.

Otra característica importante de un sistema PCM es su *resistencia a la interferencia* provocada por impulsos dispersos o diafonía. La presencia combinada del ruido del canal y de la interferencia provoca el aumento del umbral de error necesario para la operación satisfactoria del sistema PCM. Sin embargo, si se proporciona un margen adecuado sobre el umbral de error en el primer lugar, el sistema puede resistir la presencia de cantidades más o menos grandes de interferencia. Dicho de otra manera, un sistema PCM es *robusto* con respecto al ruido del canal y la interferencia.

TABLA 3.3 ***Influencia de E_b/N_0 sobre la probabilidad de error***

E_b/N_0	*Probabilidad de error* P_e	*Para una tasa de bits de* 10^5 *b/s; esto equivale aproximadamente a un error cada*
4.3 dB	10^{-2}	10^{-3} segundos
8.4 dB	10^{-4}	10^{-1} segundos
10.6 dB	10^{-6}	10 segundos
12.0 dB	10^{-8}	20 minutos
13.0 dB	10^{-10}	1 día
14.0 dB	10^{-12}	3 meses

3.9 *Multiplexado por división de tiempo*

El teorema de muestreo proporciona las bases para transmitir la información contenida en una señal de mensaje limitado en banda $m(t)$ con una secuencia de muestras $m(t)$ tomadas de manera uniforme a una frecuencia que suele ser un poco superior a la frecuencia de Nyquist. Una característica importante del proceso de muestreo es una *conservación de tiempo*. Es decir, la transmisión de las muestras de mensaje ocupan el canal de comunicación por sólo una fracción del intervalo de muestreo en un esquema periódico, y en esta forma parte del intervalo entre muestras adyacentes se desocupa para que lo usen otras fuentes de mensaje independientes en un esquema de tiempo compartido. A partir de ahí obtenemos un *sistema de multiplexado por división de tiempo* (TDM), que permite la utilización conjunta de un canal de comunicación común por una pluralidad de fuentes de mensaje independientes sin interferencia mutua entre ellas.

El concepto de TDM se ilustra por medio del diagrama de bloques de la figura 3.19. Cada señal de mensaje de entrada se restringe primero en un ancho de banda a partir de un filtro antitraslape pasobajas para eliminar las frecuencias que no son esenciales a una representación adecuada de la señal. Las salidas del filtro pasobajas se aplican luego a un *conmutador*, que suele ponerse en práctica al utilizar circuitería de conmutación electrónica. La función del conmutador es doble: 1) tomar una muestra angosta de cada uno de los N mensajes de entrada a una frecuencia f_s que es un poco mayor que $2W$, donde W es la frecuencia de corte del filtro antitraslape y 2) intercalar en forma secuencial estas N muestras dentro del intervalo de muestreo T_s. En realidad esta última función es la esencia de la operación de multiplexado por división de tiempo. Luego del proceso de conmutación, la señal multiplexada se aplica a un *modulador por pulsos*, cuyo propósito es transformar la señal multiplexada en una forma adecuada para transmitirse por el canal de comunicación. Resulta claro que el uso del multiplexado por división del tiempo introduce un factor de expansión del ancho de banda N, ya que el esquema debe apretar N muestras obtenidas de N fuentes de mensaje independiente en una franja de tiempo igual a un intervalo de muestreo. En el extremo receptor del sistema, la señal que se recibe se aplica a un *demodulador de pulsos*, que efectúa la operación inversa del modulador por pulsos. Las muestras angostas que se producen a la salida del demodulador por pulsos se distribuyen en los filtros de reconstrucción pasobajas apropiados por medio de un *deconmutador*, que opera en *sincronismo* con el conmutador en el transmisor. Esta sincronización resulta esencial para una operación satisfactoria del sistema. La manera en que esta sincronización se pone en práctica depende de modo natural del método de modulación por pulsos que se utiliza para transmitir la secuencia de muestras multiplexada.

El sistema TDM es sumamente sensible a la dispersión en el canal común; es decir, a las variaciones de amplitud con la frecuencia o a la falta de proporcionalidad de la fase con la frecuencia. Por tanto, es necesaria la ecualización exacta de las respuestas tanto en magnitud como en fase del canal para asegurar una operación satisfactoria del sistema. Este asunto se analiza en el

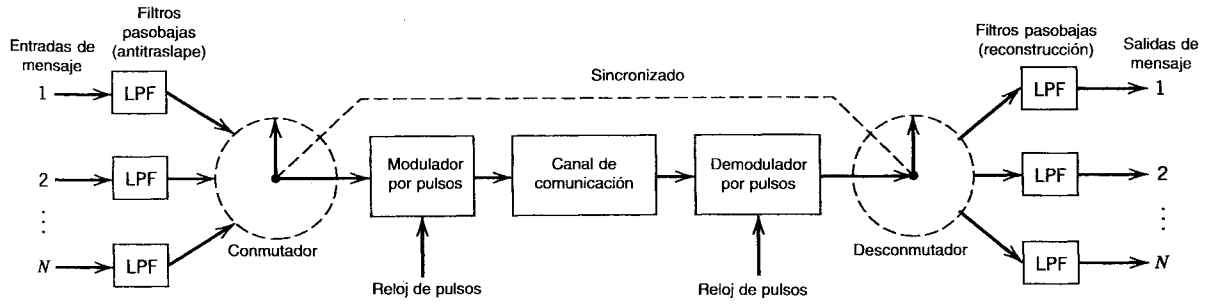


FIGURA 3.19 Diagrama de bloques del sistema TDM.

capítulo 4. Sin embargo, a diferencia de la FDM, para una aproximación de primer orden la TDM es inmune a las no linealidades en el canal como fuente de diafonía. La razón para este comportamiento es que las señales de mensaje diferentes no se aplican simultáneamente en el canal.

■ SINCRONIZACIÓN

En aplicaciones en las que se utiliza la PCM, es natural multiplexar diferentes fuentes de mensaje por medio de la división de tiempo, por lo cual cada fuente mantiene su individualidad a lo largo de la jornada del transmisor al receptor. Esta individualidad explica la facilidad comparativa con la cual las fuentes de mensaje pueden eliminarse o reinsertarse en un sistema de multiplexado por división del tiempo. Cuando aumenta el número de fuentes de mensaje, el intervalo de tiempo que es posible asignar a cada fuente tiene que reducirse, puesto que la totalidad de las mismas debe ajustarse a un intervalo de tiempo igual al recíproco de la frecuencia de muestreo, lo cual, a su vez, quiere decir que se reduce la duración permisible de una palabra de código que representa a una sola muestra. Sin embargo, los pulsos tienden a ser más difíciles de generar y transmitir cuando disminuye su duración. Además, si los pulsos se vuelven demasiado cortos, los deterioros en el medio de transmisión empiezan a interferir con la operación apropiada del sistema. En consecuencia, en la práctica es necesario restringir el número de fuentes de mensaje independientes a las que puedan incluirse dentro de un grupo de división del tiempo.

En cualquier caso, para un sistema PCM con multiplexado por división del tiempo que opere en forma satisfactoria, se vuelve necesario que las operaciones de temporización en el receptor, salvo las de pérdida de tiempo en la transmisión y en la repetición regenerativa, sigan muy de cerca las operaciones correspondientes en el transmisor. De manera general, esto se suma al requisito de un reloj local en el receptor para mantener el mismo tiempo que un reloj estándar distante en el transmisor, con la excepción de que el reloj local es un poco más lento en una cantidad correspondiente al tiempo que se requiere para transportar las señales de mensaje del transmisor al receptor. Un procedimiento posible para sincronizar los relojes del transmisor y el receptor consiste en reservar un elemento o pulso de código al final de una *trama* (consistente en una palabra de código que se obtiene de cada una de las fuentes de mensaje independientes en sucesión) y transmitir este pulso sólo en cualquier otra trama. En un caso de este tipo, el receptor incluye un circuito que buscaría el patrón de unos y ceros alternándose a la mitad de la tasa de tramas, con lo que se establecería sincronización entre el transmisor y el receptor.

Cuando se interrumpe la trayectoria de transmisión, es altamente improbable que los relojes del transmisor y el receptor continúen indicando el mismo tiempo durante un periodo prolongado. Por tanto, al llevar a cabo un proceso de sincronización, debemos fijar un procedimiento ordenado para detectar su pulso. El procedimiento consiste en observar los elementos de código uno por uno hasta que se detecta el pulso de sincronización. Es decir, después de observar un elemento de código particular lo suficiente para establecer la ausencia del pulso de sincronización, el reloj del receptor se ajusta de nuevo mediante un elemento de código y se observa el elemento de código que sigue. Este *proceso de búsqueda* se repite hasta que se detecta el pulso de sincronización. Claramente, el tiempo que se requiere para la sincronización depende del momento en el cual se restableció la transmisión apropiada.

▶ EJEMPLO 3.2 El sistema T1

En este ejemplo describimos las importantes características de un sistema PCM conocido como *sistema T1*,[6] el cual transporta 24 canales de voz por pares independientes de alambres con repetidores regenerativos espaciados aproximadamente a intervalos de 2 km. El sistema portador T1 es fundamental para la Jerarquía de Conmutación Digital Estadounidense que se explicó en la sección 3.10.

Una señal de voz (masculina o femenina) está limitada esencialmente a una banda de 300 a

3100 Hz en el sentido de que las frecuencias fuera de esta banda no contribuyen mucho a la eficiencia de la articulación. En realidad, los circuitos telefónicos que responden a este intervalo de frecuencias ofrecen un servicio bastante satisfactorio. Por tanto, es una costumbre pasar la señal de voz por un filtro pasobajas con frecuencia de corte de aproximadamente 3.1 kHz antes del muestreo. Por tanto, con $W = 3.1$ kHz, el valor nominal de la frecuencia de Nyquist es 6.2 kHz. La señal de voz filtrada suele muestrearse a una frecuencia un poco mayor, a saber, 8 kHz, que es la frecuencia de muestreo *estándar* en los sistemas electrónicos.

Para la compresión y la expansión, el sistema T1 utiliza una característica *lineal por secciones* consistente en 15 segmentos lineales para aproximar la ley μ logarítmica para la ecuación (3.48) con la constante $\mu = 255$. Esta aproximación se construye de tal manera que los puntos finales del segmento se encuentran en la curva de compresión calculada a partir de la ecuación (3.48) y sus proyecciones sobre el eje vertical están uniformemente espaciadas. La tabla 3.4 indica las proyecciones de los puntos finales del segmento sobre el eje principal y el tamaño de los segmentos individuales. La tabla se normaliza hasta 8159, de modo que todos los valores se representan como números enteros. El segmento cero de la aproximación es un segmento colineal, que pasa por el origen; contiene un total de 30 niveles de decisión uniformes. Los segmentos lineales $1a$, $2a$, . . . , $7a$ se encuentran sobre el eje horizontal, en tanto que los segmentos lineales $1b$, $2b$, . . . , $7b$ yacen debajo del eje horizontal; cada uno de estos 14 segmentos contiene 16 niveles de decisión uniformes. Para un segmento colineal cero, los niveles de decisión a la entrada del cuantizador son ± 1, ± 3, . . . , ± 31, y los niveles de representación correspondientes a la salida del cuantizador son 0, ± 1, . . . , ± 15. En cuanto a los segmentos lineales $1a$ y $1b$, los niveles de decisión a la entrada del cuantizador son $\pm$ 31, ± 35, . . . , ± 95, y los niveles de representación correspondientes a la salida del cuantizador son ± 16, ± 17, . . . , ± 31, y así sucesivamente para los otros segmentos lineales.

Hay un total de $31 + (14 \times 16) = 255$ niveles de representación asociados con la característica de compresión-expansión de 15 segmentos que se describió antes. Para alojar este número de niveles de representación, cada uno de los 24 canales de voz utiliza un código binario con una palabra de 8 bits. El primer bit indica si la muestra de voz de entrada es positiva o negativa; este bit es un 1 si es positivo y un 0 si es negativo. Los siguientes 3 bits de la palabra de código identifican el segmento particular dentro del cual se encuentra la amplitud de la muestra de voz de entrada, y los últimos 4 bits identifican el nivel de representación real dentro de ese segmento.

Con una frecuencia de muestreo de 8 kHz, cada trama de la señal multiplexada ocupa un periodo de 125 μs. En particular, ésta consiste en 24 palabras de 8 bits, más un bit único que se agrega al final de la trama para fines de sincronización. Por tanto, cada trama consta de un total de $(24 \times 8) + 1 = 193$ bits. De modo correspondiente, la duración de cada bit es igual a 0.647 μs, y la velocidad de transmisión que resulta es igual a 1.544 megabits por segundo (Mb/s).

Además de la señal de voz, un sistema telefónico debe dejar pasar también señales de supervisión especiales hasta el extremo remoto. Esta *información de la transmisión de señal* se necesita para

TABLA 3.4 ***La característica del compresor-expansor de 15 segmentos ($\mu = 255$)***

Número de segmentos lineales	*Tamaño del escalón*	*Proyecciones de los puntos extremos de los segmentos sobre el eje horizontal*
0	2	± 31
$1a$, $1b$	4	± 95
$2a$, $2b$	8	± 223
$3a$, $3b$	16	± 479
$4a$, $4b$	32	± 991
$5a$, $5b$	64	± 2015
$6a$, $6b$	128	± 4063
$7a$, $7b$	256	± 8159

transmitir pulsos de marcado, así como señales telefónicas de colgar/descolgar. En el sistema T1 este requerimiento se realiza de la manera siguiente. Cada 6 tramas, el bit menos significativo (esto es, el octavo) de cada canal de voz se retarda y un *bit de transmisión de señal* se inserta en su lugar, produciendo de esa manera una operación de $7\frac{5}{6}$ bits promedio por cada entrada de voz. La secuencia de bits de transmisión se transmite de esa manera a una frecuencia igual a la tasa de muestreo de 8 kHz dividida entre seis; esto es, 1.333 kb/s. Esta frecuencia de transmisión de señal se aplica a cada uno de los 24 canales de entrada.

3.10 *Multiplexores digitales*

En la sección 3.9 presentamos la idea del multiplexado por división del tiempo mediante el cual un grupo de señales analógicas (por ejemplo, señales de voz) se muestrean en forma secuencial en el tiempo a una frecuencia de muestreo *común* y después se aplica el multiplexado para la transmisión por una línea común. En esta sección consideramos el multiplexado de señales digitales a diferentes tasas de bit. Esto nos permite combinar varias señales digitales, como salidas de computadora, señales de voz digitalizadas, facsímil digitalizado y señales de televisión, en una única secuencia de datos (a una tasa de bits considerablemente superior que cualquiera de las entradas). La figura 3.20 muestra un diagrama conceptual de la operación *digital de multiplexado-demultiplexado.*

El multiplexado de señales digitales se realiza utilizando un *procedimiento de intercalamiento bit por bit* con un interruptor selector que toma en forma secuencial un bit de cada línea entrante y luego lo aplica a la línea común de alta velocidad. En el extremo receptor del sistema, la salida de esta línea común se separa en sus componentes individuales de baja velocidad y luego se entrega a sus destinos respectivos.

Los multiplexores digitales se clasifican en dos grupos principales. Uno se usa para adquirir secuencias de datos de tasa de bits relativamente baja que se originan de computadoras digitales y realizan el multiplexado para la transmisión TDM por la red pública telefónica conmutable. La puesta en práctica de este primer grupo de multiplexores requiere el uso de módems (moduladores-demoduladores), que se explican en el capítulo 6.

Un segundo grupo de multiplexores digitales forma parte del servicio de transmisión de datos que ofrecen transmisores de telecomunicación tales como AT&T. En particular, estos multiplexores constituyen una *jerarquía digital* que realiza el multiplexado por división de tiempo de secuencias de baja tasa de bits y las convierte en secuencias de tasas de bits superiores. Los detalles de las tasas de bits que se acomodan en la jerarquía varían de un país a otro. Sin embargo, un rasgo mundial de la jerarquía es que se inicia a 64 kb/s, lo cual corresponde a la representación de la PCM estándar de una señal de voz. Una secuencia de bits entrante a esta tasa, sin que importe su origen, recibe el nombre de *cero de señal digital (DS0).* En *Estados Unidos, Canadá* y *Japón*[7], la jerarquía se apega a la de la TDM digital estadounidense, según se describe a continuación:

- La *jerarquía de primer nivel* combina 24 secuencias de bits DS0 para obtener una *señal digital única (DS1)* a 1.544 Mb/s, la cual se transporta sobre el sistema T1 descrito en el

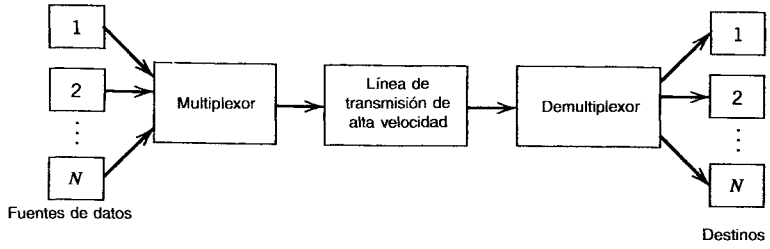


FIGURA 3.20 Diagrama conceptual del multiplexado-demultiplexado.

ejemplo 3.2. Estas secuencias de bit se denominan la *velocidad principal* en la jerarquía digital, debido a que es la velocidad de bit más baja que existe fuera del conmutador digital. El *conmutador digital* es un dispositivo compuesto por memoria y lógica, cuya función es meramente la conmutación de las señales digitales, de ahí el nombre.

- El *multiplexor de segundo nivel* combina cuatro secuencias de bit DS1 para obtener una *señal digital dos (DS2)* a 6.312 Mb/s.
- El *multiplexor de tercer nivel* combina siete secuencias de bits DS2 para obtener una *señal digital tres (DS3)* a 44.736 Mb/s.
- El *multiplexor de cuarto nivel* combina seis secuencias de bits DS3 para obtener una *señal digital cuarta (DS4)* a 274.176 Mb/s.
- El *multiplexor de quinto nivel,* el final en la jerarquía, combina dos secuencias de bits de DS4 para obtener una *señal digital cinco (DS5)* a 560.160 Mb/s.

Advierta que la tasa de bits de una señal digital producida por uno de estos multiplexores es un poco superior que el múltiplo preestablecido de la tasa de bit entrante debido al apiñamiento de bits incorporado en el diseño de cada multiplexor; dicho apiñamiento se analizará más adelante.

Además, es importante admitir que las funciones de un equipo de transmisión digital implican meramente transportar una secuencia de bits sin importar qué es lo que estos últimos significan. Sin embargo, los conmutadores digitales en los dos extremos de la instalación deben tener una comprensión común de cómo interpretar los bits dentro de la secuencia, ya sea que representen voz o datos, formatos de tramas, formatos de transmisión, etcétera.

Hay algunos problemas básicos implicados en el diseño de un multiplexor digital, independientemente de su agrupamiento.

1. Las señales digitales no pueden intercalarse directamente en un formato que posibilite su separación al final, a menos que sus tasas de bits se ajusten a un reloj común. En vez de eso, deben tomarse medidas para realizar la *sincronización* de las señales digitales de entrada, de manera que sea posible intercalarlas en forma apropiada.
2. La señal multiplexada debe incluir alguna forma de tramas, de manera que sus componentes individuales puedan identificarse en el receptor.
3. El multiplexor tiene que manejar pequeñas variaciones en las tasas de bits de las señales digitales entrantes. Por ejemplo, un cable coaxial de 1000 km que conduce 3×10^8 pulsos por segundo tendrá en tránsito cerca de un millón de pulsos, con cada uno de ellos ocupando casi un metro del cable. Una variación de 0.01% en el retraso de la propagación, producto de una disminución de 1°F de temperatura, originará cien pulsos menos en el cable. Es claro que el multiplexor debe absorber estos pulsos.

Para adaptar los requerimientos de sincronización y el ajuste de velocidad a fin de acomodar pequeñas variaciones en las velocidades de los datos de entrada, es posible recurrir a una técnica conocida como *apiñamiento de bits.* La idea aquí es tener la tasa de bits que sale del multiplexor un poco mayor que la suma de las tasas de bits máximas esperadas de los canales de entrada al apiñar pulsos adicionales que no contienen información. Todas las señales digitales de entrada se apiñan con un número de bits suficiente para elevar cada una de sus tasas de bits e igualar la del reloj que se genera localmente. Para conseguir el apiñamiento de bits, cada señal digital entrante o secuencia de bits se alimenta en un *almacén elástico* en el multiplexor. Dicho almacén es un dispositivo que almacena una secuencia de bits de tal manera que ésta puede leerse externamente a una tasa diferente a la cual se lee internamente. En el demultiplexor, los bits apiñados deben evidentemente eliminarse de la señal multiplexada. Lo anterior requiere un método que es posible utilizar para identificar los bits apiñados. Para ilustrar un método de este tipo, y para presentar también un método que ofrece sincronización de trama, describimos el formato de señal de *multiplexor M12* de AT&T, el cual está diseñado para combinar cuatro se-

cuencias de bit DS1 en una secuencia de bits DS2. Éste es el segundo nivel de la jerarquía digital que se explicó antes.

EJEMPLO 3.3 Formato de señal del multiplexor M12 de AT&T

La figura 3.21 ilustra el formato de señal del multiplexor M12. Cada trama se divide en cuatro subtramas. Se transmite la primera (línea inicial en la figura 3.21), luego la segunda, la tercera y la cuarta, en ese orden.

La intercalación bit por bit de las cuatro secuencias de bit DS1 entrantes se usa para acumular un total de 48 bits, 12 de cada entrada. Después se inserta un *bit de control* mediante el multiplexor. Cada trama contiene un total de 24 bits de control, separados por secuencias de 48 bits de datos. Se utilizan tres tipos de bits de control en el multiplexor M12 para proporcionar sincronización y la indicación de trama, así como para identificar cuál de las cuatro señales de entrada se ha apiñado. Estos bits de control se designan con las letras F, M y C en la figura 3.12. Sus funciones son las siguientes:

1. Los bits de control F, dos por subtrama, constituyen los pulsos de trama *principales*. Los subíndices en los bits de control F denotan el bit real (0 o 1) transmitido. De tal manera, la secuencia de trama principal es $F_0F_1F_0F_1F_0F_1F_0F_1$ o 01010101.
2. Los bits de control M, uno por subtrama, forman pulsos de trama *secundarios*, para identificar las cuatro subtramas. En este caso también los subíndices de control M denotan el bit real (0 o 1) transmitido. De ese modo, la secuencia de trama secundaria es $M_0M_1M_1M_1$ o 0111.
3. Los bits de control C, tres por subtrama, son *indicadores de apiñamiento*. En particular, C_I se refiere al canal de entrada I, C_{II} se refiere al canal de entrada II, etc. Por ejemplo, los tres bits de control C en la primera subtrama que sigue a M_0 en la primera subtrama son indicadores de apiñamiento para la primera secuencia de bits DS1. La inserción de un bit de apiñamiento en esta cadena de bits DS1 se indica igualando a uno los tres bits de control C. Para indicar que no hay apiñamiento, los tres se igualan a 0. Si los tres bits de control C indican apiñamiento, el bit de apiñamiento se ubica en la posición del primer bit de información asociado con la primera secuencia de bits DS1 que sigue al bit de control F_1 en la misma subtrama. De manera similar, las secuencias de bit DS1 segunda, tercera y cuarta pueden apiñarse, según se requiera. Utilizando *decodificación lógica de mayoría* en el receptor, es posible detectar un error único en cualquiera de los tres bits de control C. Esta forma de decodificación indica simplemente que la mayoría de los bits de control C determinan si se transmitió una secuencia exclusiva de unos o exclusiva de ceros. De tal modo, los tres unos o combinaciones de dos unos y un cero indican que está presente un bit de apiñamiento en la secuencia de información, siguiendo al bit de control F_1 en la subtrama pertinente. Por otro lado, tres ceros o combinaciones de dos ceros y un 1 indican que no se está utilizando apiñamiento.

El demultiplexor en la unidad M12 de recepción busca primero la secuencia de trama principal $F_0F_1F_0F_1F_0F_1F_0F_1$. De ese modo establece la identidad para las cuatro secuencias de bit DS1 de entrada y también para los bits de control M y C. A partir de la secuencia $M_0M_1M_1M_1$ se verifica la trama correcta de bits de control C. Por último, las cuatro secuencias de bits DS1 se demultiplexan y desapiñan en forma apropiada.

El formato de señal que acaba de describirse tiene dos elementos de seguridad:

1. Es posible, aunque improbable, que con apenas la secuencia $F_0F_1F_0F_1F_0F_1F_0F_1$, una de las secuencias de bit DS1 entrante pueda contener una secuencia similar. Esto podría provocar en este caso

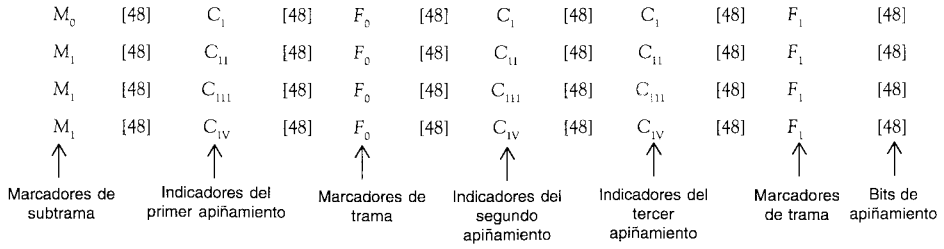


FIGURA 3.21 Formato de señales del multiplexor M12 de AT&T.

que el receptor se cerrara en la secuencia equivocada. La presencia de la secuencia $M_0M_1M_1M_1$ proporciona la verificación de la secuencia genuina $F_0F_1F_0F_1F_0F_1F_0F_1$, asegurando de ese modo que las cuatro secuencias de bits DS1 se están demultiplexando en forma apropiada.

2. La capacidad de corrección de errores únicos incorporada en los bits de control C asegura el desapiñamiento adecuado de las cuatro secuencias de bits DS1.

La capacidad del multiplexor M12 para ajustarse a pequeñas variaciones en las velocidades de los datos de entrada puede calcularse a partir del formato de la figura 3.21. En cada trama M, definida como el intervalo que contiene un ciclo de bits $M_0M_1M_1M_1$, un bit puede apiñarse en cada una de las cuatro secuencias de bits de entrada DS1. Cada señal de ese tipo tiene $12 \times 6 \times 4 = 288$ posiciones en cada trama M. Asimismo, la secuencia de bits T1 tiene una tasa de bits igual a 1.544 Mb/s. Por tanto, cada entrada puede incrementarse en

$$1.544 \times 10^3 \times \frac{1}{288} = 5.4 \text{ kb/s}$$

Este resultado es mucho mayor que el cambio esperado en la tasa de bits de la secuencia de bits entrantes DS1. Por tanto, se concluye que el uso de sólo un bit de apiñamiento por canal de entrada en cada trama resulta suficiente para ajustarse a las variaciones esperadas en la velocidad de entrada.

El reloj local que determina la tasa saliente de bits determina asimismo la *velocidad de apiñamiento* nominal S, que se define como el número promedio de bits apiñados por canal en cualquier trama. El multiplexor M12 se diseña para $S = 1/3$. En consecuencia, la tasa de bits nominal de la secuencia de bits DS2 es

$$1.544 \times 4 \times \frac{49}{48} \times \frac{288}{288 - S} = 6.312 \text{ Mb/s}$$

En esta forma asegura también que la frecuencia nominal de reloj DS2 es un múltiplo de 8 kHz (la frecuencia de muestreo nominal de una señal de voz), que es una característica deseable. ◀

3.11 *Virtudes, limitaciones y modificaciones de la PCM*

En un sentido general, la modulación por codificación de pulsos (PCM) ha surgido como el esquema de modulación más favorable en la transmisión de señales que contienen información analógica como las de video y voz. Las ventajas de la PCM posiblemente puedan calcarse del uso de *la codificación de pulsos para la representación digital de señales analógicas*, una característica que las distingue de todos los demás métodos de modulación analógica. Podemos resumir las importantes ventajas de la PCM del modo siguiente:

1. *Robustez* al ruido y la interferencia del canal.
2. *Regeneración* eficiente de la señal codificada a lo largo de la trayectoria de transmisión.
3. *Intercambio* eficiente del ancho de banda incrementado del canal por una relación señal a ruido mejorada, obedeciendo a una ley exponencial.
4. Un *formato uniforme* para la transmisión de diferentes tipos de señales de banda base, y de ahí su integración con otras formas de datos digitales con una red común.
5. *Facilidad* comparativa con la cual las fuentes de mensaje tal vez se eliminen o reincorporen en un sistema de multiplexado por división de tiempo.
6. Comunicación *segura* mediante el uso de esquemas de modulación o encriptado especiales; el encriptado y el desencriptado de datos se explican en el apéndice 5.

Sin embargo, tales ventajas se alcanzan a costa de una mayor complejidad del sistema y al incrementar el ancho de banda del canal. Estos dos aspectos se considerarán a su vez más adelante.

Si bien el uso de la PCM implica muchas operaciones complejas, en la actualidad todas ellas pueden ponerse en práctica a un costo conveniente utilizando chips *a muy grande escala de integración* (VLSI) disponibles comercialmente y/o hechos bajo pedido. En otras palabras, ya está cubierto el requisito de la tecnología de dispositivos para la ejecución de un sistema PCM. Asimismo, con la mejora continua en la tecnología VLSI, es probable que veamos un uso cada vez mayor de la PCM en la transmisión digital de señales analógicas.

Sin embargo, si un requisito necesario es la simplicidad de la puesta en práctica, entonces es necesario recurrir a la modulación delta como una alternativa para la modulación por codificación de pulsos. En la modulación delta, la señal de banda base se "sobremuestrea" de manera intencional para posibilitar el uso de una estrategia de cuantización simple en la construcción de la señal codificadora; la modulación delta se explica en la sección 3.12.

Volviendo al aspecto del ancho de banda, admitimos que el aumento en el requisito del ancho de banda de la PCM quizá haya tenido una razón por una preocupación justificable en el pasado. Sin embargo, hoy en día no constituye una preocupación real por dos razones diferentes. En primer lugar, la disponibilidad creciente de *canales de comunicación de banda ancha* significa que el ancho de banda ya no es una restricción del sistema en la manera tradicional que solía serlo. La liberación de la restricción del ancho de banda ha sido posible por el despliegue de los satélites de comunicación para la transmisión y el uso cada vez mayor de fibras ópticas para la conexión en redes; una discusión relativa a estos conceptos del canal de comunicación se presentó en el capítulo de antecedentes y preliminares.

La segunda razón estriba en que a través del uso de técnicas complejas de *compresión de datos* es en realidad posible eliminar la redundancia presente en forma inherente en una señal PCM y reducir de ese modo la tasa de bits de los datos transmitidos sin una degradación seria en el desempeño del sistema. En realidad, la mayor complejidad de procesamiento (y, por tanto, el mayor costo de la puesta en práctica) se intercambia por una tasa de bits reducida y, en consecuencia, un requerimiento de ancho de banda menor. Una motivación importante para la reducción de tasa de bits corresponde a la comunicación segura por canales de radio que son inherentemente de baja capacidad.

3.12 *Modulación delta*

En la *modulación delta*[8] (DM), se sobremuestrea una señal de mensaje entrante (es decir, a una frecuencia mucho mayor que la de Nyquist) para aumentar de modo intencional la correlación entre muestras adyacentes de la señal. De esta forma se permite el uso de la estrategia de cuantización simple para construir la señal codificada.

En su forma básica, la DM proporciona una *aproximación de escalera* a la versión sobremuestreada de la señal de mensaje, como se ilustra en la figura 3.22*a*. La diferencia entre la entrada y la aproximación se cuantiza sólo en dos niveles, a saber, $\pm\Delta$, correspondiente a diferencias positivas y negativas. De ese modo, si la aproximación cae por debajo de la señal en cualquier momento de muestreo, se incrementa en Δ. Si, por otro lado, la aproximación se ubica sobre la señal, ésta se disminuye en Δ. Siempre y cuando la señal no cambie demasiado rápido de muestra a muestra, encontramos que la aproximación de escalera se mantiene dentro de $\pm\ \Delta$ de la señal de entrada.

Dejemos que $m(t)$ denote la señal de entrada (mensaje) y $m_q(t)$ su aproximación de escalera. Por conveniencia en la presentación, adoptamos la presente notación que se usa de modo muy común en la literatura relativa al procesamiento de señales digitales:

$$m[n] = m(nT_s), \qquad n = 0, \pm 1, \pm 2, \ldots$$

donde T_s es el periodo de muestreo y $m(nT_s)$ es una muestra de la señal $m(t)$ tomada al tiempo $t = nT_s$, y de modo similar para las muestras de otras señales en tiempo continuo. En esas

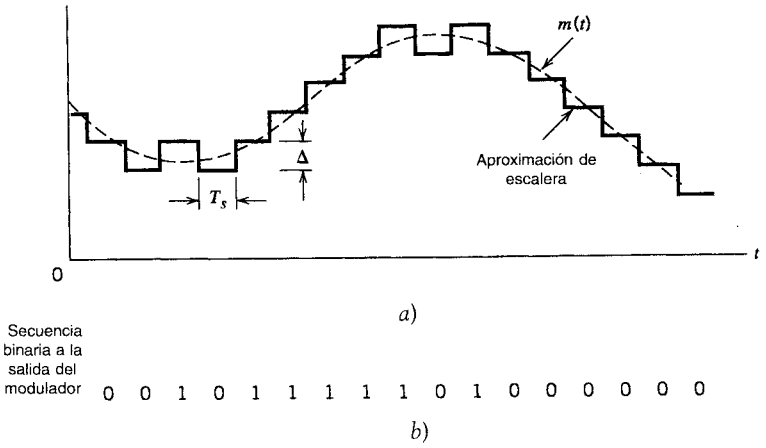


FIGURA 3.22 Ilustración de la modulación delta.

condiciones es posible formalizar los principios básicos de la modulación delta en el siguiente conjunto de relaciones en tiempo discreto:

$$e[n] = m[n] - m_q[n-1] \tag{3.52}$$

$$e_q = \Delta \ \text{sgn}(e[n]) \tag{3.53}$$

$$m_q[n] = m_q[n-1] + e_q[n] \tag{3.54}$$

donde $e[n]$ es una *señal de error* que representa la diferencia entre la muestra presente $m[n]$ de la señal de entrada y la última aproximación $m_q[n-1]$ a ella, $e_q[n]$ es la versión cuantizada de $e[n]$ y sgn(·) es la función signo. Por último, la salida del cuantizador $m_q[n]$ se codifica para producir la señal de DM.

La figura 3.22*a* ilustra la forma en la cual la aproximación de escalera $m_q(t)$ sigue las variaciones de la señal de entrada $m(t)$ de acuerdo con las ecuaciones (3.52)-(3.54), y la figura 3.22*b* exhibe la correspondiente secuencia binaria a la salida del modulador delta. Es claro que en el sistema de modulación delta la frecuencia de información de transmisión es simplemente igual a la frecuencia de muestreo $f_s = 1/T_s$.

La simplicidad es la virtud principal de la modulación delta. Ésta puede generarse al aplicar la versión muestreada de la señal de mensaje entrante a un modulador que incluye un *comparador, cuantizador* y un *acumulador* interconectados, como se indica en la figura 3.23*a*. El block marcado como z^{-1} dentro del acumulador representa un *retardo unitario*, esto es, un retardo igual a un periodo de muestreo. (La variable z suele usarse en la transformada z, la cual resulta básica para el análisis de señales y sistemas en tiempo discreto.) Los detalles del modulador siguen directamente de las ecuaciones (3.52)–(3.54). El comparador calcula la diferencia entre sus dos entradas. El cuantizador consiste en un *limitador rígido* con una relación de entrada-salida que es una versión escalada de la función signo. La salida del cuantizador se aplica entonces a un acumulador, produciendo el resultado

$$\begin{aligned} m_q[n] &= \Delta \sum_{i=1}^{n} \text{sgn}(e[i]) \\ &= \sum_{i=1}^{n} e_q[i] \end{aligned} \tag{3.55}$$

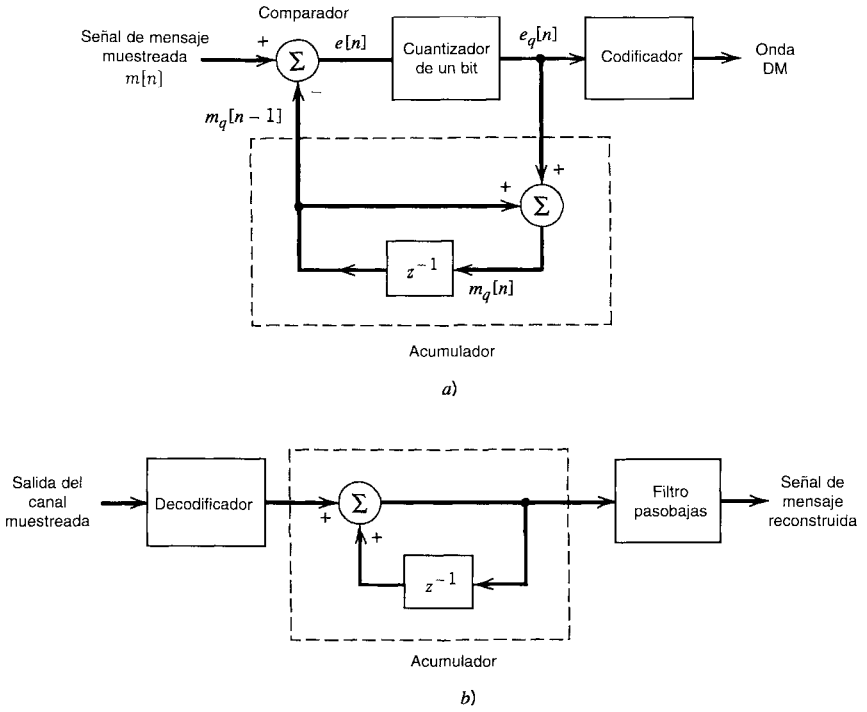


FIGURA 3.23 Sistema DM. *a*) Transmisor. *b*) Receptor.

que se obtiene resolviendo las ecuaciones (3.53) y (3.54) para $m_q[n]$. Por tanto, en el instante de muestreo nT_s, el acumulador aumenta la aproximación en un escalón Δ en una dirección positiva o negativa, según el signo algebraico de la muestra de error $e[n]$. Si la muestra de entrada $m[n]$ es mayor que la aproximación más reciente $m_q[n]$, se aplica un incremento positivo $+\Delta$ a la aproximación. En cambio, si la muestra de entrada es más pequeña, se aplica un incremento negativo $-\Delta$ a la aproximación. En esta forma, el acumulador hace lo mejor que puede para seguir las muestras de entrada de escalón en escalón (de amplitud $+\Delta$ o $-\Delta$). En el receptor que se muestra en la figura 3.23*b* se reconstruye la aproximación de escalera $m_q(t)$ al hacer pasar la secuencia de pulsos positivos y negativos, producidos a la salida del decodificador, a través de un acumulador de una manera similar a la que se usó en el transmisor. El ruido de cuantización fuera de banda en la forma de onda de escalera de alta frecuencia $m_q(t)$ se rechaza al hacerlo pasar por un filtro pasobajas, como en la figura 3.23*b*, con un ancho de banda igual al ancho de banda del mensaje original.

La modulación delta está sujeta a dos tipos de errores de cuantización: distorsión de sobrecarga de pendiente y ruido granular. Explicaremos primero el caso de la distorsión de sobrecarga de pendiente.

Observamos que la ecuación (3.54) es el equivalente digital de la integración en el sentido de que representa la acumulación de incrementos positivos y negativos de magnitud Δ. Además, denotando el error de cuantización por $q[n]$, de acuerdo con

$$m_q[n] = m[n] + q[n] \tag{3.56}$$

observamos de la ecuación (3.52) que la entrada al cuantizador es

$$e[n] = m[n] - m[n-1] + q[n-1] \tag{3.57}$$

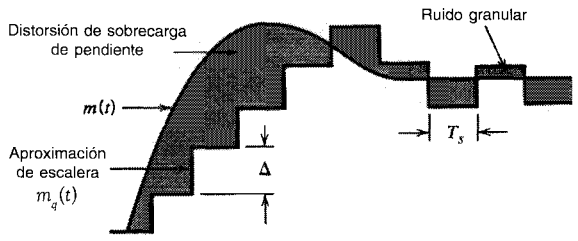


FIGURA 3.24 Ilustración de las dos formas diferentes del error de cuantización en la modulación delta.

De tal modo, excepto por el error $q[n-1]$, la entrada del cuantizador es una *primera diferencia hacia atrás* de la señal de entrada, que puede considerarse como una aproximación digital a la derivada de la señal de entrada o, de manera equivalente, como la inversa de un proceso de integración digital. Si consideramos la pendiente máxima de la forma de entrada original $m(t)$, resulta claro que para aumentar la secuencia de muestras $\{m_q[n]\}$ tan rápido como la secuencia de entrada de muestras $\{m[n]\}$ en una región de pendiente máxima de $m(t)$, requerimos que la condición

$$\frac{\Delta}{T_s} \geq \text{máx}\left|\frac{dm(t)}{dt}\right| \tag{3.58}$$

se satisfaga. De otro modo encontramos que el tamaño de escalón Δ es demasiado pequeño para que la aproximación de escalera $m_q(t)$ siga un segmento inclinado de la forma de entrada $m(t)$, con el resultado de que $m_q(t)$ cae detrás de $m(t)$, como se ilustra en la figura 3.24. Esta condición se conoce como *sobrecarga de pendiente,* y el error de cuantización resultante recibe el nombre de *distorsión (ruido) de sobrecarga de pendiente.* Advierta que en vista de que la pendiente máxima de la aproximación de escalera $m_q(t)$ está fija para un tamaño de escalón Δ, los aumentos y decrementos en $m_q(t)$ tienden a ocurrir a lo largo de líneas rectas. Por esta razón, un modulador Δ que usa un tamaño de escalón fijo a menudo se conoce como *modulador delta lineal.*

En contraste con la distorsión de sobrecarga de pendiente, el *ruido granular* ocurre cuando el tamaño de escalón Δ es demasiado grande en relación con la característica de pendiente local de la forma de onda de entrada $m(t)$, causando por ello que la aproximación de escalera $m_q(t)$ persiga de cerca un segmento relativamente plano de la forma de onda de entrada; este fenómeno se ilustra también en la figura 3.24. El ruido granular es análogo al ruido de cuantización en un sistema PCM.

De este modo vemos que existe la necesidad de un gran tamaño de escalón para alojar un amplio intervalo dinámico, en tanto que se requiere un tamaño de escalón pequeño para una representación exacta de las señales de nivel relativamente bajo. Por tanto, resulta claro que la elección del tamaño de escalón óptimo que minimice el valor cuadrático medio del error de cuantización en un modulador delta lineal será el resultado de un compromiso entre la distorsión de sobrecarga de pendiente y el ruido granular. Para satisfacer un requisito de este tipo necesitamos hacer "adaptable" al modulador delta, a causa de que el tamaño de escalón se hace variar de acuerdo con la señal de entrada; este asunto se explica adicionalmente en el experimento de computadora que se presenta en la sección 3.16.

MODULACIÓN DELTA-SIGMA

Como se señaló antes, la entrada del cuantizador en la forma convencional en la modulación delta puede considerarse como una aproximación a la *derivada* de la señal de mensaje entrante. Este comportamiento conduce a una desventaja de la modulación delta en el sentido de que perturbaciones de la transmisión como el ruido originan un error acumulativo en la señal demodulada. Esta

desventaja puede superarse *integrando* la señal de mensaje antes de la modulación delta. El uso de la integración del modo que se describe aquí tiene también los siguientes efectos benéficos:

- Se preenfatiza el contenido de baja frecuencia de la señal de entrada.
- Se incrementa la correlación entre muestras adyacentes de la entrada del modulador delta, lo que tiende a mejorar el desempeño de todo el sistema al reducir la varianza de la señal de error a la entrada del cuantizador.
- Se simplifica el diseño del receptor.

Un esquema de modulación delta que incorpore integración en su entrada se denomina *modulación delta-sigma* (D-ΣM).[9] Para ser más precisos, esto debe llamarse *modulación sigma-delta*, ya que la integración se efectúa en realidad antes que la modulación delta. A pesar de eso, la primera terminología es la que suele emplearse en la literatura.

La figura 3.25*a* muestra el diagrama de bloques de un sistema de modulación delta-sigma; en este diagrama, la señal de mensaje $m(t)$ se define en su forma en tiempo continuo, lo que significa que el modulador por pulso consiste ahora en un limitador rígido seguido por un multiplicador; el último componente se alimenta también a partir de un generador de pulsos externos (reloj) para producir una señal codificada de 1 bit. El uso de integración a la entrada del transmisor requiere claramente de una acentuación de señal inversa, a saber, de diferenciación, en el receptor. Sin embargo, se elimina la necesidad de esta diferenciación debido a que la misma se cancela por la integración en el

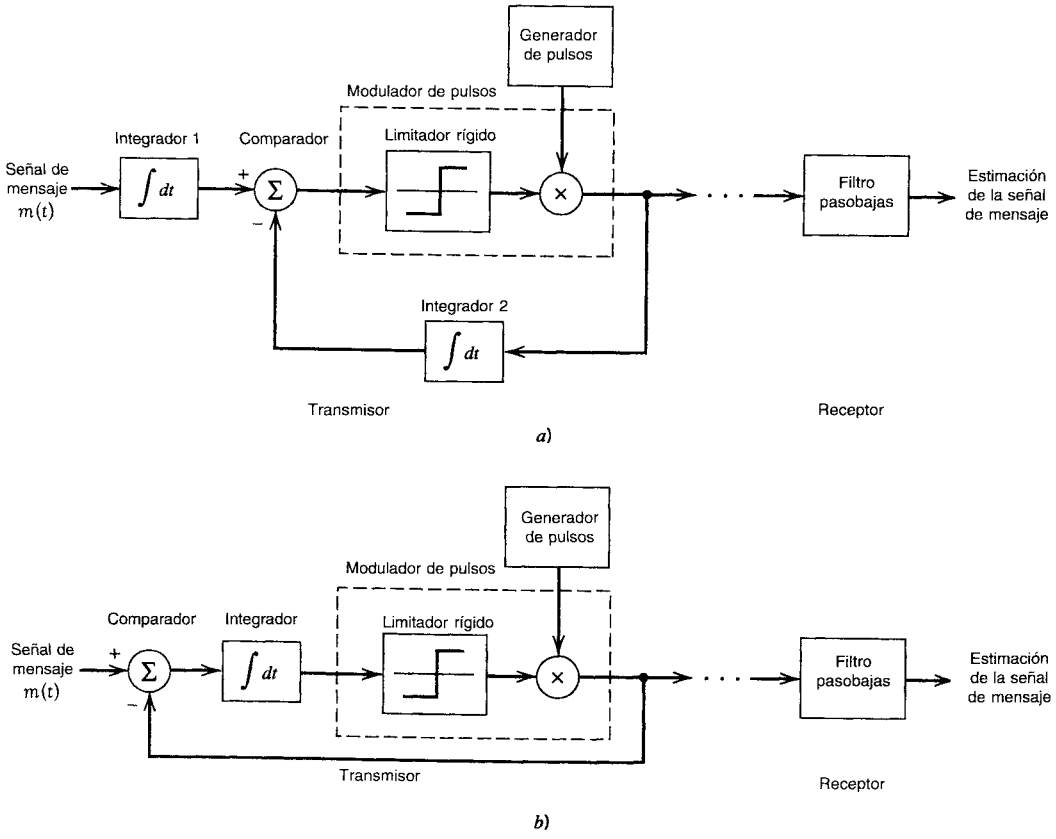


FIGURA 3.25 Dos versiones equivalentes del sistema de modulación delta-sigma.

receptor DM convencional. De este modo, el receptor de un sistema de modulación delta-sigma consiste simplemente en un filtro pasobajas, como se indica en la figura 3.25*a*.

Asimismo, notamos que la integración es básicamente una operación lineal, por lo que es posible simplificar el diseño del transmisor combinando los dos integradores 1 y 2 de la figura 3.25*a* en un solo integrador situado después del comparador, como se indica en la figura 3.25*b*. Esta última forma de sistema de modulación delta-sigma no sólo es más simple que el de la figura 3.25*a*, sino que también proporciona una interpretación interesante de la modulación delta-sigma como una versión "uniformada" de una modulación por codificación de pulsos de un bit: el término *uniformidad* se refiere a que la salida del comparador se integra antes de la cuantización, y el término de *un bit* reenuncia tan solo que el cuantizador está compuesto por un limitador rígido con sólo dos niveles de representación.

En la modulación delta, la simplicidad de las puestas en práctica tanto del transmisor como del receptor se consigue al utilizar una frecuencia de muestreo bastante mayor que la necesaria en la modulación por codificación de pulsos. El precio que se paga por este beneficio es un aumento correspondiente en la transmisión y, por tanto, en el ancho de banda del canal. Sin embargo, hay aplicaciones donde el ancho de banda del canal es difícil de conseguir, en cuyo caso tenemos el requisito opuesto al de la modulación delta. Específicamente, tal vez nos interese intercambiar la mayor complejidad del sistema por un ancho de banda de canal reducido. Una operación básica de procesamiento de señales para conseguir este último objetivo de diseño es la predicción, cuya forma lineal se presenta a continuación.

3.13 *Predicción lineal*

Considere un *filtro en tiempo discreto de respuesta al impulso de duración finita (FIR)* configurado como en la figura 3.26, el cual implica el uso de tres bloques funcionales:

1. Un conjunto de p elementos de retraso unitario, cada uno de los cuales se representa por medio de z^{-1}.
2. Un conjunto de multiplicadores que implican los coeficientes de filtro $w_1, w_2, \ldots, w_p$.
3. Un conjunto de "sumadores" que se utiliza para sumar las versiones escaladas de las entradas retrasadas $x[n-1], x[n-2], \ldots, x[n-p]$ para producir la salida $\hat{x}[n]$. La salida del filtro $\hat{x}[n]$ o, con más exactitud, la *predicción lineal* de la entrada, se define consecuentemente mediante la *suma de convolución*

$$\hat{x}[n] = \sum_{k=1}^{p} w_k x[n-k] \tag{3.59}$$

donde p, el número de elementos de retraso unitario, recibe el nombre de *orden de predicción*.

La muestra real en el tiempo nT_s es $x[n]$. El *error de predicción*, denotado por $e[n]$, se define como la diferencia entre $x[n]$ y la predicción $\hat{x}[n]$, como se indica mediante

$$e[n] = x[n] - \hat{x}[n] \tag{3.60}$$

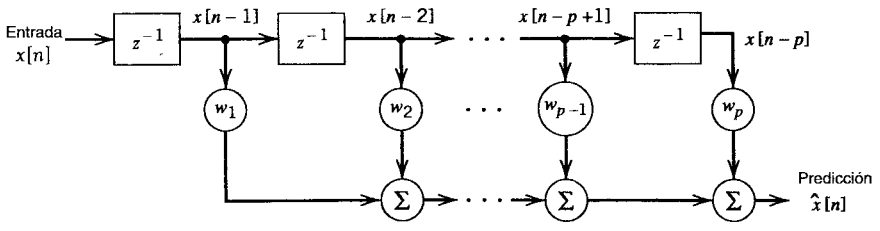


FIGURA 3.26 Diagrama de bloques de un filtro de predicción lineal de orden p.

El objetivo de diseño es elegir los coeficientes del filtro $w_1, w_2, \ldots, w_p$ de modo que se minimice un *índice de desempeño, J*, definido como el error cuadrático medio:

$$J = E[e^2[n]] \tag{3.61}$$

Al sustituir las ecuaciones (3.59) y (3.60) en la (3.61) y al expandir los términos, es posible reformular el índice de desempeño como

$$J = E[x^2[n]] - 2\sum_{k=1}^{p} w_k E[x[n]x[n-k]] + \sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k E[x[n-j]x[n-k]] \tag{3.62}$$

Suponemos que la señal de entrada $x(t)$ es la función que se muestra de un proceso estacionario $X(t)$ de media cero; esto es, $E[x[n]]$ es cero para toda n. Defina

$$\begin{aligned}
\sigma_X^2 &= \text{varianza de una muestra del proceso } X(t) \text{ en el tiempo } nT_s \\
&= E[x^2[n]] - (E[x[n]])^2 \\
&= E[x^2[n]] \\
R_X(kT_s) &= \text{autocorrelación del proceso } X(t) \text{ para el retardo de } kT_s \\
&= R_X[k] \\
&= E[x[n]x[n-k]]
\end{aligned}$$

En consecuencia, es posible reescribir la ecuación (3.62) en la forma simplificada

$$J = \sigma_X^2 - 2\sum_{k=1}^{p} w_k R_X[k] + \sum_{j=1}^{p}\sum_{k=1}^{p} w_j w_k R_X[k-j] \tag{3.63}$$

Por tanto, al diferenciar el índice de desempeño J con respecto al coeficiente del filtro w_k, igualando el resultado a cero y rearreglando después los términos, obtenemos

$$\sum_{j=1}^{p} w_1 R_X[k-j] = R_X[k] = R_X[-k], \qquad k = 1, 2, \ldots, p \tag{3.64}$$

Las expresiones de optimización (3.64) se denominan *ecuaciones Wiener-Hopf* para la predicción lineal.

Encontramos conveniente reformular las ecuaciones de Wiener-Hopf (3.64) en forma de matriz. Sea

$$\begin{aligned}
\mathbf{w}_o &= \text{vector de coeficientes óptimos de } p \text{ por } 1 \\
&= [w_1, w_2, \ldots, w_p]^T \\
\mathbf{r}_X &= \text{vector de autocorrelación de } p \text{ por } 1 \\
&= [R_X[1], R_X[2], \ldots, R_X[p]]^T \\
\mathbf{R}_X &= \text{matriz de autocorrelación de } p \text{ por } 1 \\
&= \begin{bmatrix}
R_X[0] & R_X[1] & \cdots & R_X[p-1] \\
R_X[1] & R_X[0] & \cdots & R_X[p-2] \\
\vdots & \vdots & & \vdots \\
R_X[p-1] & R_X[p-2] & \cdots & R_X[0]
\end{bmatrix}
\end{aligned}$$

De tal modo podemos simplificar el conjunto de ecuaciones (3.64) como

$$\mathbf{R}_X \mathbf{w}_o = \mathbf{r}_X \tag{3.65}$$

Suponemos que la matriz de autocorrelación $\mathbf{R}_X$ es *no singular,* de modo que existe su inversa. En ese caso podemos resumir la ecuación (3.65) para el vector de coeficientes $\mathbf{w}_o$ multiplicando ambos miembros de esta ecuación por la *matriz inversa* $\mathbf{R}_X^{-1}$, obteniendo la solución óptima

$$\mathbf{w}_o = \mathbf{R}_X^{-1}\mathbf{r}_X \tag{3.66}$$

Advierta que todos los elementos en la diagonal principal en la matriz de autocorrelación $\mathbf{R}_X$ son iguales a $R_X[0] = \sigma_X^2$, y que los elementos de cualquier otra diagonal paralela a la diagonal principal son también iguales. Una matriz cuadrada que tenga esta cualidad se dice que será de *Toeplitz,* lo cual es una consecuencia directa de la suposición de que la señal de entrada $x(t)$ se toma de un proceso estacionario. El significado práctico de la propiedad de Toeplitz es que la matriz de correlación $\mathbf{R}_X$ está definida unívocamente por el conjunto de valores de autocorrelación $R_X[0], R_X[1], \ldots, R_X[p-1]$. El vector de correlación $\mathbf{r}_X$ se define mediante el conjunto de valores de autocorrelación $R_X[1], R_X[2], \ldots, R_X[p]$. Por tanto, se concluye que los coeficientes del filtro p del predictor óptimo están definidos en forma única por la varianza $\sigma_X^2 = R_X[0]$ y los valores p de la función de autocorrelación del proceso $X(t)$ para retrasos de $T_s, 2T_s, \ldots, pT_s$.

El valor cuadrático medio mínimo del error de predicción se obtiene al sustituir la ecuación (3.64) en la (3.63), lo cual origina (después de la simplificación)

$$J_{\text{mín}} = \sigma_X^2 - \mathbf{r}_X^T\mathbf{R}_X^{-1}\mathbf{r}_X \tag{3.67}$$

El término cuadrático $\mathbf{r}_X^T\mathbf{R}_X^{-1}\mathbf{r}_X$ es positivo siempre. En consecuencia, el error cuadrático medio $J_{\text{mín}}$ del predictor lineal óptimo definido por la ecuación (3.67) siempre es menor que la varianza σ_X^2 de la muestra de entrada que se está pronosticando.

Predicción adaptable lineal

El empleo de la ecuación (3.66) para calcular el vector de peso de una predicción lineal requiere del conocimiento de la función de autocorrelación $R_X[k]$ de la secuencia de entrada $\{x[n]\}$ para retardos $k = 0, 1, \ldots, p$, donde p es el orden de predicción. ¿Qué sucede si no se cuenta con el conocimiento de $R_X[k]$ para k variable? En estas situaciones, que ocurren con frecuencia en la práctica, podemos recurrir al uso de un *predictor adaptable.*

El predictor es adaptable en el siguiente sentido:

- El cálculo de los pesos en derivación w_k, $k = 1, 2, \ldots, p$, procede de una manera "recursiva", empezando a partir de algunos valores iniciales arbitrarios de los pesos de las derivaciones.
- El algoritmo que se utiliza para ajustar los pesos de las derivaciones (de una iteración a la siguiente) está "autodiseñado", y opera sólo con base en los datos disponibles.

El objetivo del algoritmo es determinar el punto mínimo de la *superficie de error* en forma de tazón que describe la dependencia de la función de costo J sobre los pesos de las derivaciones. Por tanto, es intuitivamente razonable que los ajustes sucesivos a los pesos de las derivaciones del pronosticador se realicen en la dirección del descenso más pronunciado de la superficie de error, es decir, en una dirección opuesta al *vector gradiente* cuyos elementos se definen mediante

$$g_k = \frac{\partial J}{\partial w_k}, \quad k = 1, 2, \ldots, p \tag{3.68}$$

Ésta es en realidad la idea detrás del *método del descenso más pronunciado.* Considere que $w_k[n]$ denota el valor del k-ésimo peso de derivación en la iteración n. Entonces el valor actualizado de este peso en la iteración $n + 1$ está definido por

$$w_k[n+1] = w_k[n] - \frac{1}{2}\mu g_k, \quad k = 1, 2, \ldots, p \tag{3.69}$$

donde μ es el *parámetro del tamaño de escalón* que controla la velocidad de adaptación, y el factor 1/2 se incluye por conveniencia en la presentación. Al diferenciar la función de costo J de la ecuación (3.63) con respecto a w_k, encontramos de inmediato que

$$\begin{aligned} g_k &= -2R_x[k] + 2\sum_{j=1}^{p} w_j R_x[k-j] \\ &= -2E[x[n]x[n-k]] + 2\sum_{j=1}^{p} w_j E[x[n-j]x[n-k]], \quad k = 1, 2, \ldots, p \end{aligned} \tag{3.70}$$

Esta fórmula para g_k podría tener una simplificación adicional, lo cual se logra utilizando *valores instantáneos como estimaciones* de las funciones de autocorrelación $R_x[k]$ y $R_x[k-j]$. Esto es, ignoramos los operadores de esperanza en la ecuación (3.70) para facilitar el proceso adaptable en un esquema paso por paso. De ese modo es posible expresar la estimación correspondiente de g_k en la iteración n como

$$\hat{g}_k[n] = -2x[n]x[n-k] + 2\sum_{j=1}^{p} w_j[n]x[n-j]x[n-k], \quad k = 1,2,\ldots,p \tag{3.71}$$

Advierta que para una entrada $x[n]$ tomada de un proceso estacionario, el gradiente g_k es una cantidad determinística, en tanto que la estimación $\hat{g}_k[n]$ es el valor de muestreo de una variable aleatoria.

En cualquier caso, al sustituir la ecuación (3.71) en la (3.69) y factorizar el término común $x[n-k]$, es posible escribir

$$\begin{aligned} \hat{w}_k[n+1] &= \hat{w}_k[n] + \mu x[n-k]\left(x[n] - \sum_{j=1}^{p} \hat{w}_j[n]x[n-j]\right) \\ &= \hat{w}_k[n] + \mu x[n-k]e[n], \quad k = 1, 2, \ldots, p \end{aligned} \tag{3.72}$$

donde $e[n]$ es el *error de predicción* definido como

$$e[n] = x[n] - \sum_{j=1}^{p} \hat{w}_j[n]x[n-j] \tag{3.73}$$

En las ecuaciones (3.72) y (3.73) hemos utilizado $\hat{w}_k$ como una *estimación* del k-ésimo peso de derivación para distinguirlo del valor real w_k. Advierta además que $x[n]$ desempeña el papel de una "respuesta deseada" para calcular los ajustes recursivos aplicados a los pesos de derivación del predictor.

Las ecuaciones (3.72) y (3.73) constituyen el popular *algoritmo de mínimos cuadrados* para la predicción adaptable lineal, cuya operación se describe en la figura 3.27. La razón de la popularidad de este algoritmo de filtrado adaptable es la simplicidad de su puesta en práctica. En particular, la complejidad computacional del algoritmo, medido en términos de sumas y multiplicaciones, es *lineal* en el orden de predicción p.

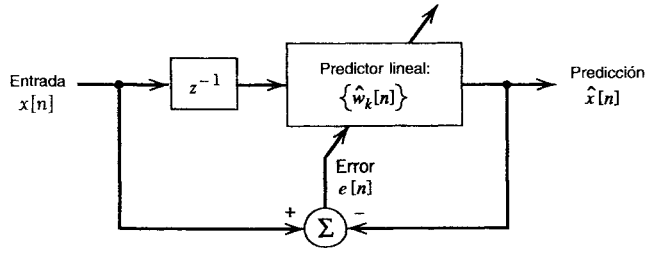


FIGURA 3.27 Diagrama de bloques que ilustra el proceso de predicción lineal adaptable.

El algoritmo de mínimos cuadrados es uno de filtrado adaptable *estocástico*; es estocástico a causa de que, partiendo de la *condición inicial* definida por $\{w_k[0]\}_{k=1}^{p}$, intenta encontrar el punto mínimo de la superficie del error siguiendo una trayectoria en zigzag. Además, nunca encuentra su punto mínimo con exactitud. Más bien, ejecuta un movimiento al azar alrededor del punto mínimo de la superficie de error, una vez que se establecen las condiciones de estado estable.

Con la disposición de este material acerca de la predicción lineal, estamos listos para explicar las mejoras prácticas en el desempeño de la modulación por codificación de pulsos.

3.14 *Modulación diferencial por codificación de pulsos*

Cuando se muestrea una señal de voz o video a una frecuencia un poco mayor que la de Nyquist, como suele efectuarse en la modulación por codificación de pulsos, la señal muestreada que se origina exhibe un alto grado de correlación entre muestras adyacentes. El significado de esta alta correlación es que, en un sentido promedio, la señal no cambia rápidamente de una muestra a la siguiente y, en consecuencia, la diferencia entre muestras adyacentes tiene una varianza que es más pequeña que la de la propia señal. Cuando se codifican estas muestras altamente correlacionadas, como en el sistema PCM estándar, la señal codificada que se produce contiene *información redundante*, lo cual quiere decir que los símbolos que no son absolutamente esenciales para la transmisión de información se generan como resultado del proceso de codificación. Al eliminar esta redundancia antes de la codificación, obtenemos una señal codificada más eficiente, lo cual es la idea básica detrás de la modulación diferencial por codificación de pulsos.

Ahora bien, si conocemos el comportamiento pasado de una señal hasta cierto punto en el tiempo, es posible utilizar la predicción para estimar un valor futuro de la señal, como se describe en la sección 3.13. Supóngase entonces que la señal de banda base $m(t)$ se muestrea a una velocidad $f_s = 1/T_s$ para producir la secuencia $\{m[n]\}$ cuyas muestras están separadas T_s segundos. El hecho de que es posible predecir valores futuros de la señal $m(t)$ proporciona la motivación para el esquema de *cuantización diferencial* que se presenta en la figura 3.28*a*. En dicho esquema, la señal de entrada para el cuantizador se define mediante

$$e[n] = m[n] - \hat{m}[n] \tag{3.74}$$

que es la diferencia entre la muestra de entrada sin cuantizar $m[n]$ y una predicción de ella, denotada por $\hat{m}[n]$. Este valor predicho se produce utilizando un filtro de predicción lineal cuya entrada, como veremos, consiste en una versión cuantizada de la muestra de entrada $m[n]$. La señal de diferencia $e[n]$ es el error de predicción, ya que es la cantidad mediante la cual el filtro de predicción falla al predecir exactamente la entrada. Codificando la salida del cuantizador, como en la figura 3.28*a*, obtenemos una variante de la PCM conocida como *modulación diferencial por codificación de pulsos*[10] (DPCM).

La salida del cuantizador puede expresarse como

$$e_q[n] = e[n] + q[n] \tag{3.75}$$

donde $q[n]$ es el error de cuantización. De acuerdo con la figura 3.28*a*, la salida del cuantizador $e_q[n]$ se agrega al valor predicho $\hat{m}[n]$ para producir la entrada del filtro de predicción.

$$m_q[n] = \hat{m}[n] + e_q[n] \tag{3.76}$$

Al sustituir la ecuación (3.75) en la (3.76), obtenemos

$$m_q[n] = \hat{m}[n] + e[n] + q[n] \tag{3.77}$$

Sin embargo, a partir de la ecuación (3.74) observamos que el término $\hat{m}[n] + e[n]$ es igual a la muestra de entrada $m[n]$. Por tanto, podemos simplificar la ecuación (3.77) como

$$m_q[n] = m[n] + q[n] \tag{3.78}$$

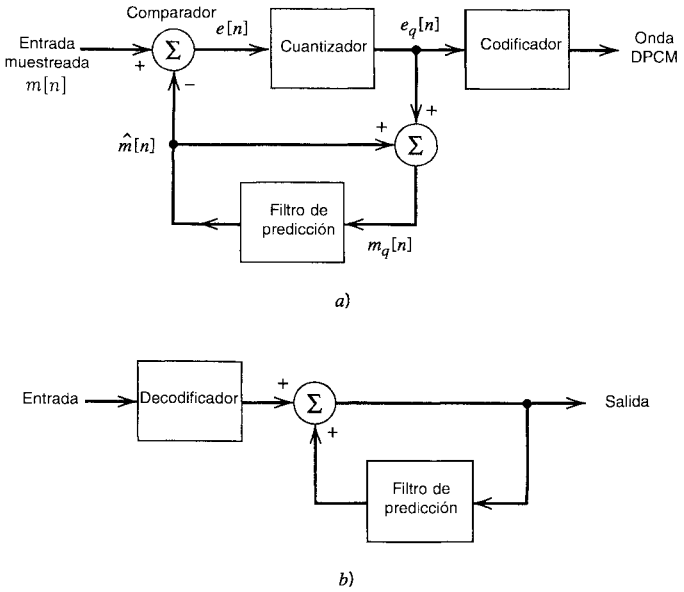


FIGURA 3.28 Sistema DPCM. *a*) Transmisor. *b*) Receptor.

que representa una versión cuantizada de la muestra de entrada $m[n]$. Esto es, sin que importen las propiedades del filtro de predicción, la muestra cuantizada $m_q[n]$ en la entrada del filtro de predicción difiere de la muestra de entrada original $m[n]$ por el error de cuantización $q[n]$. En consecuencia, si la predicción es buena, la varianza del error de predicción $e[n]$ será más pequeña que la varianza $m[n]$, de modo que un cuantizador con un número determinado de niveles puede ajustarse para producir un error de cuantización con una varianza más pequeña que la que sería posible si la muestra de entrada $m[n]$ se cuantizara directamente como en un sistema PCM estándar.

El receptor para reconstruir la recepción cuantizada de la entrada se muestra en la figura 3.28*b*. Está compuesto por un decodificador para reconstruir la señal de error cuantizado. La versión cuantizada de la entrada original se reconstruye a partir de la salida del decodificador utilizando el mismo filtro de predicción que se empleó en el transmisor de la figura 3.28*a*. A falta de ruido del canal, encontramos que la señal codificada de la entrada del receptor es idéntica a la señal codificada en la salida del transmisor. Por tanto, la correspondiente salida del receptor es igual a $m_q[n]$, la cual difiere de la entrada original $m[n]$ sólo por el error de cuantización $q[n]$ en el que se incurre como resultado de la cuantización del error de predicción $e[n]$.

Del análisis precedente observamos que, en un ambiente sin ruido, los filtros de predicción en el transmisor y el receptor operan sobre la misma secuencia de muestras, $m_q[n]$. Con este propósito en mente se agrega una trayectoria de realimentación al cuantizador en el transmisor, como se muestra en la figura 3.28*a*.

La modulación diferencial por codificación de pulsos incluye la modulación delta como un caso especial. En particular, al comparar el sistema DPCM de la figura 3.28 con el sistema DM de la figura 3.23, vemos que éstos son básicamente similares, salvo por dos diferencias importantes: el uso de un cuantizador de un bit (nivel 2) en el modulador delta y la sustitución del filtro de predicción por un único elemento de retardo (es decir, orden de predicción cero). En términos simples, la DM es la versión de 1 bit de un DPCM. Advierta que a diferencia de un sistema PCM estándar, los transmisores tanto de DPCM como de PCM incluyen el uso de *realimentación*.

La DPCM, al igual que la DM, está sujeta a una distorsión de sobrecarga de pendiente siempre que la señal de entrada cambie demasiado con la suficiente rapidez como para que el filtro de predicción pueda seguirla. Además, al igual que la PCM, la DPCM está expuesta al ruido de cuantización.

GANANCIA DEL PROCESAMIENTO

La relación señal a ruido de salida del sistema DPCM que se muestra en la figura 3.28 es, por definición,

$$(\text{SNR})_O = \frac{\sigma_M^2}{\sigma_Q^2} \tag{3.79}$$

donde σ_M^2 es la varianza de la muestra de entrada original $m[n]$, supuesta de media cero, y σ_Q^2 es la varianza del error de cuantización $q[n]$. Es posible reescribir la ecuación (3.79) como el producto de dos factores en la forma siguiente:

$$\begin{aligned}(\text{SNR})_O &= \left(\frac{\sigma_M^2}{\sigma_E^2}\right)\left(\frac{\sigma_E^2}{\sigma_Q^2}\right) \\ &= G_p(\text{SNR})_Q\end{aligned} \tag{3.80}$$

donde σ_E^2 es la varianza del error de predicción. El factor $(\text{SNR})_Q$ es la *relación señal a ruido de cuantización*, que se define por medio de

$$(\text{SNR})_Q = \frac{\sigma_E^2}{\sigma_Q^2} \tag{3.81}$$

El otro factor G_p es la *ganancia de procesamiento* producida por el esquema de cuantización diferencial; ésta se define por medio de

$$G_p = \frac{\sigma_M^2}{\sigma_E^2} \tag{3.82}$$

La cantidad G_p, cuando es mayor que la unidad, representa una ganancia en la relación señal a ruido que se debe al esquema de cuantización diferencial de la figura 3.28. Ahora bien, para una señal de banda base (mensaje) determinada, la varianza σ_M^2 es fija, de modo que G_p se hace máxima al minimizar la varianza σ_E^2 del error de predicción $e[n]$. Por tanto, nuestro objetivo debe enfocarse en diseñar el filtro de predicción de manera que se minimice σ_E^2.

En el caso de señales de voz, se encuentra que la ventaja óptima de señal a ruido de cuantización de la DPCM sobre la PCM estándar se ubica en la vecindad de 4 a 11 dB. La mayor mejora ocurre al ir de ninguna predicción a la de primer orden, con alguna ganancia adicional que proviene del aumento del orden del filtro de predicción hasta 4 o 5, después de lo cual se obtiene poca ganancia adicional. Puesto que 6 dB de ruido de cuantización equivalen a un bit por muestra en virtud de la ecuación (3.35), la ventaja de la DPCM puede expresarse también en términos de la tasa de bits. Para una relación constante de señal a ruido de cuantización, y suponiendo una frecuencia de muestreo de 8 kHz, es posible que el empleo de la DPCM proporcione un ahorro de 8 a 16 kb/s (es decir, de 1 a 2 bits por muestra) en comparación con la PCM estándar.

3.15 *Modulación diferencial adaptable por codificación de pulsos*

El uso de la PCM para la codificación del habla a una velocidad estándar de 64 kb/s demanda un amplio ancho de banda de canal para su transmisión. Sin embargo, en ciertas aplicaciones, como la transmisión segura por canales de radio que son inherentemente de baja capacidad, el ancho de banda del canal es difícil de conseguir. En aplicaciones de este tipo existe una necesidad definida para la *codificación del habla a bajas tasas de bits, mientras se mantenga una fidelidad aceptable o calidad de reproducción.*

En la codificación del habla a bajas tasas de bits, se optimiza un codificador de onda de configuración preestablecida explotando tanto la *caracterización estadística de las formas de onda del habla* como las *propiedades de la audición*. En particular, el criterio de diseño tiene dos objetivos en mente:

1. Eliminar las redundancias de la señal de voz lo más posible.
2. Asignar los bits disponibles para codificar las partes más redundantes de la señal del habla de una manera eficiente desde la perspectiva perceptual.

Cuando procuramos reducir la tasa de bits de 64 kb/s (utilizada en la PCM estándar) a 32, 16, 8 y 4 kb/s, los esquemas que se utilizan para la eliminación de la redundancia y la asignación de bits se vuelven cada vez más complejos. Como una regla práctica, en el intervalo de 64 a 8 kb/s, la complejidad computacional (medida en términos de operaciones de multiplicación-suma) que se requiere para codificar el habla aumenta en un orden de magnitud cuando la tasa de bits se divide a la mitad, para aproximadamente la misma calidad de voz.

En esta sección describimos la *modulación diferencial adaptable por codificación de pulsos* (ADPCM),[11] la cual permite la codificación del habla a 32 kb/s mediante el uso combinado de *cuantización adaptable* y *predicción adaptable*; el número de 8 bits por muestra que se requiere en la PCM estándar se reduce de ese modo a 4. El término *adaptable* quiere decir que se responde al nivel variable y al espectro de la señal de voz de entrada. La variación del desempeño con los hablantes y del material del habla, junto con las variaciones en el nivel de señal inherente en los procesos de comunicación del habla, demandan el empleo combinado de la cuantización y la predicción adaptables para alcanzar el mejor desempeño con respecto a una amplia gama de hablantes y de situaciones en las que se efectúa dicho fenómeno.

La cuantización adaptable se refiere a un cuantizador que opera con un tamaño de escalón *variable en el tiempo* $\Delta[n]$. En cualquier instante de muestreo identificado por el índice n, se supone que el cuantizador adaptable tiene una característica de transferencia uniforme. El tamaño del escalón $\Delta[n]$ se varía hasta igualar la varianza σ_M^2 de la muestra de entrada $m[n]$. En particular, escribimos

$$\Delta[n] = \phi\hat{\sigma}_M[n] \tag{3.83}$$

donde ϕ es una constante y $\hat{\sigma}_M[n]$ es una *estimación* de la desviación estándar $\sigma_M[n]$ (es decir, la raíz cuadrada de la varianza σ_M^2). Para una entrada no estacionaria $\sigma_M[n]$ es variable en el tiempo. El problema de la cuantización adaptable de acuerdo con la ecuación (3.83) corresponde, por tanto, a calcular continuamente la estimación $\hat{\sigma}_M[n]$.

La ejecución de la ecuación (3.83) puede proceder de dos formas:

1. *Cuantización adaptable con estimación directa* (AQF), en la cual se usan muestras no cuantizadas de la señal de entrada para obtener estimaciones directas de $\sigma_M[n]$.
2. *Cuantización adaptable con estimación inversa* (AQB), en la cual se recurre a muestras de la salida del cuantizador para obtener estimaciones inversas de $\sigma_M[n]$.

El esquema AQF requiere el empleo de un *búfer* para almacenar muestras no cuantizadas de la señal de entrada del habla necesarias en el periodo de aprendizaje. Requiere también la transmisión explícita de información de nivel (por lo común, alrededor de 5 a 6 bits por muestra del tamaño del escalón) para un decodificador remoto, agobiando por ello al sistema con *información secundaria* adicional que tiene que transmitirse al receptor. Además se produce un *retraso* de procesamiento (del orden de 16 ms en el caso del habla) en la operación de codificado por el uso de la AQF, lo cual resulta inaceptable en algunas aplicaciones. Los problemas de transmisión del nivel, el uso de búfers y el retardo intrínseco de la AQF se evitan con la AQB. En este último esquema se usa la historia reciente de la salida del cuantizador para extraer información con la que se calculará el tamaño del escalón $\Delta[n]$. Por tanto, en la práctica suele preferirse la AQB en lugar de la AQF.

La figura 3.29 muestra el diagrama de bloques de un cuantizador adaptable con estimación inversa. Éste representa un sistema de realimentación no lineal; por tanto, no es evidente

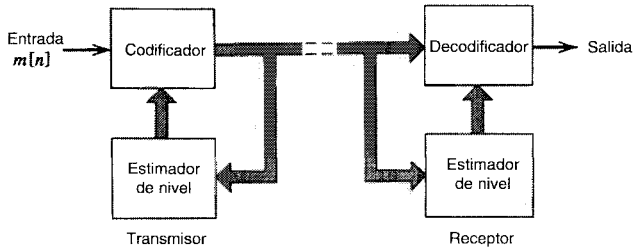


FIGURA 3.29 Cuantización adaptable con estimación inversa (AQB).

que el sistema será estable. Sin embargo, si se *acota* la entrada $m[n]$ del cuantizador, entonces también lo estarán la estimación inversa $\hat{\sigma}_M[n]$ y el tamaño de escalón correspondiente $\Delta[n]$; en estas condiciones, el sistema en realidad es *estable*.

El uso de la predicción adaptable en la ADPCM se justifica debido a que las señales del habla son inherentemente *no estacionarias*, un fenómeno que se manifiesta por sí solo en el hecho de que la función de autocorrelación y la densidad espectral de potencia de las señales de voz son funciones variables en el tiempo de sus respectivos argumentos. Esto implica que el diseño de predictores para tales entradas deba ser similarmente variable en el tiempo, es decir, adaptable. Como en el caso de la cuantización adaptable, hay dos esquemas para efectuar la predicción adaptable:

1. *Predicción adaptable con estimación directa* (APF), en la cual se utilizan muestras no cuantizadas de la señal de entrada para obtener estimaciones de los coeficientes del predictor.
2. *Predicción adaptable con estimación inversa* (APB), en la cual se emplean muestras de la salida del cuantizador y el error de predicción para obtener estimaciones de los coeficientes del predictor.

Sin embargo, la APF está sujeta a las mismas desventajas intrínsecas (información secundaria, uso de búfers y retardo) que la AQF. Estas desventajas se eliminan al utilizar el esquema APB que se muestra en la figura 3.30, donde la caja marcada "lógica para predicción adaptable" representa al algoritmo para actualizar los coeficientes del predictor. En el último esquema, los coeficientes óptimos del predictor se estiman con base en los datos cuantizados y transmitidos. Por tanto, pueden actualizarse tan frecuentemente como se desee, digamos, de muestra a muestra. Por tanto, la APB es el método de predicción que se prefiere para la ADPCM.

El algoritmo de mínimos cuadrados para el predictor, descrito en la sección 3.13, y un esquema adaptable para el cuantizador, fundamentado en la ecuación (3.83), se han combina-

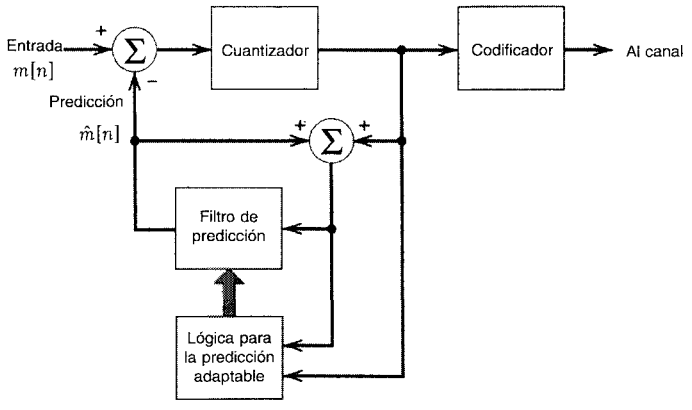


FIGURA 3.30 Predicción adaptable con estimación inversa (APB).

do de un modo sincrónico en el diseño tanto del codificador como del decodificador. El desempeño de esta combinación es tan eficaz a 32 kb/s que en la actualidad la ADPCM se acepta internacionalmente como una técnica de codificación estándar para las señales de voz, junto con 64 kb/s utilizando la PCM estándar.

3.16 *Experimento de computadora: modulación delta adaptable*

Una forma simple de la AQB se encuentra en la modificación de una modulación delta lineal (LDM) para formar la *modulación delta adaptable* (ADM). Son dos los principios que fundamentan a todos los algoritmos ADM:

1. Si los errores sucesivos son de polaridad opuesta, entonces el modulador delta está operando en su modo granular; en este caso, quizá sea ventajoso reducir el tamaño del escalón.
2. Sin embargo, si los errores sucesivos son de la misma polaridad, entonces el modulador delta está operando en su modo de sobrecarga de pendiente; en ese segundo caso, es necesario aumentar el tamaño del escalón.

De modo que al variar el tamaño del escalón de acuerdo con este principio, se permite que el modulador delta haga frente a los cambios en la señal de entrada.

La figura 3.31 muestra el diagrama de bloques de una ADM basada en el aumento o decremento del tamaño de escalón por un factor de 50% en cada iteración del proceso adaptable. El algoritmo para la adaptación del tamaño del escalón está definido por[12]

$$\Delta[n]=\begin{cases}\dfrac{|\Delta[n-1]|}{m_q[n]}\left(m_q[n]+0.5m_q[n-1]\right) & \text{si } \Delta[n-1]\geq \Delta_{\text{mín}} \\ \Delta_{\text{mín}} & \text{si } \Delta[n-1]< \Delta_{\text{mín}}\end{cases} \tag{3.84}$$

donde $\Delta[n]$ es el tamaño del escalón en la iteración (escalón de tiempo) n del algoritmo y $m_q[n]$ es la salida del cuantizador de 1 bit que es igual a $\pm$ 1.

En este experimento utilizamos una señal de entrada senoidal para demostrar el desempeño de reconstrucción del logaritmo ADM basado en la ecuación (3.84), y lo comparamos con el desempeño de un modulador delta lineal (LDM) correspondiente. Los detalles del experimento son los siguientes:

Señal de entrada:

$$m(t) = A \operatorname{sen}(2\pi f_m t)$$

donde la amplitud $A = 10$, la frecuencia $f_m = f_s/100$ y f_s = frecuencia de muestreo.

Modulación delta lineal (LDM):

$$\text{Tamaño del escalón } \Delta[n]=1 \text{ para toda } n$$

Modulación delta adaptable (ADM):

$$\Delta_{\text{mín}} = \frac{1}{8}$$

Los resultados del experimento se grafican en la figura 3.32. La parte *a* de esta figura corresponde a la LDM, y la parte *b* a la ADM. A partir de las formas de onda que se presentan aquí, es posible señalar las siguientes observaciones:

- La ADM *sigue* los cambios en la señal de entrada senoidal mucho mejor que la LDM. Esta mejora en el desempeño de la ADM se debe a la adaptación del tamaño del escalón en las

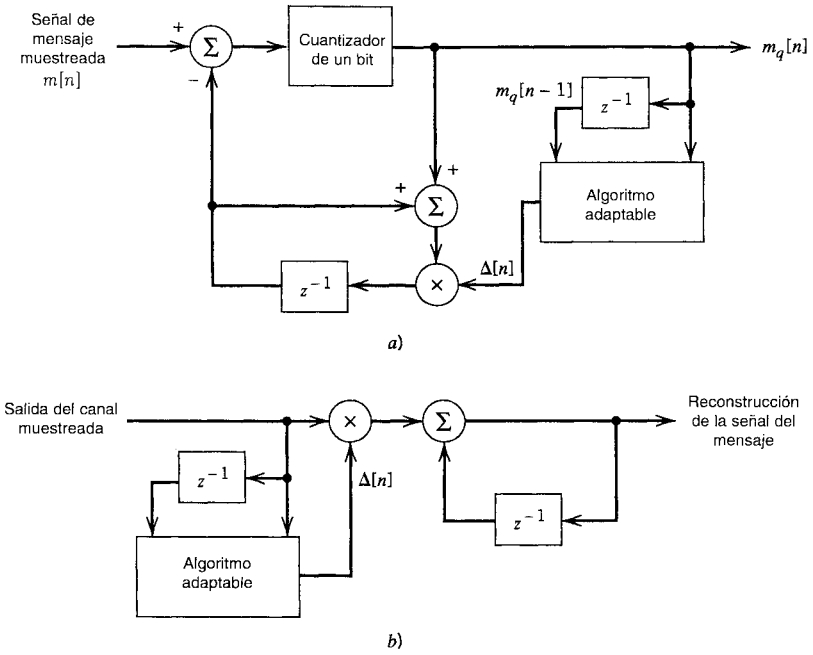


FIGURA 3.31 Sistema de modulación delta adaptable: *a*) Transmisor. *b*) Receptor.

iteraciones sucesivas del algoritmo. En particular, el tamaño reducido de escalón de la ADM origina errores de cuantización cercanos a las extremidades de la señal de entrada más pequeños que en el caso de la LDM. Sin embargo, ambos esquemas de modulación dan lugar a errores de cuantización comparables en regiones de la señal de entrada donde la pendiente es moderadamente elevada.

- El mejor desempeño de seguimiento de la ADM produce una señal de salida con una tasa de bits mucho menor, en promedio, que la LDM.

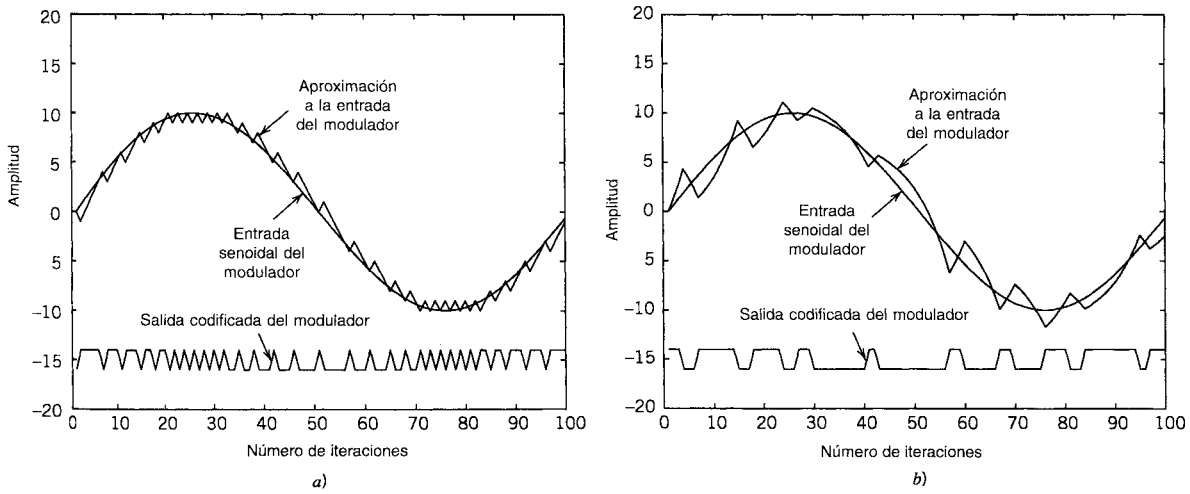


FIGURA 3.32 Formas de onda que resultan del experimento de computadora sobre modulación delta: *a*) Modulación delta lineal. *b*) Modulación delta adaptable.

3.17 *Estándar de codificación de audio MPEG*

Las señales del habla (voz) y de audio son similares a causa de que, en ambos casos, la calidad del esquema de codificación se sustenta en las propiedades de la percepción auditiva humana. En el caso de las señales del habla, tenemos esquemas de codificación eficientes (por ejemplo, la ADPCM) debido a que se dispone de un modelo de producción del habla. Desafortunadamente, no existe nada similar para las señales de audio.

En esta sección volvemos a considerar el *estándar de codificación de audio* MPEG-1 que se describió brevemente en el capítulo de antecedentes y preliminares; MPEG corresponde a las siglas en inglés de Motion Picture Experts Group, y el sufijo 1 da a entender que es el primero de una serie de varios estándares.[13] Al igual que la ADPCM, el estándar de codificación de audio MPEG-1 es un sistema de compresión con pérdida, aunque difiere de aquél en un importante aspecto práctico: el estándar MPEG-1 es capaz de conseguir una *compresión transparente y perceptualmente sin pérdidas* de señales de audio estereofónicas a elevadas frecuencias de muestreo. En particular, las pruebas de audición subjetivas efectuadas por el comité de audio de la MPEG en condiciones de audición muy difíciles, han demostrado que incluso con una razón de compresión de 6 a 1, las señales de audio codificada y original son perceptualmente indistinguibles.

El estándar de codificación de audio MPEG-1 alcanza este desempeño notable al explotar dos características psicoacústicas del sistema auditivo humano:

1. *Bandas críticas.*
El oído interno[14] del sistema auditivo representa el espectro de potencia de señales entrantes sobre una escala no lineal en la forma de bandas de frecuencia limitadas denominadas *bandas críticas.* La banda de frecuencia audible, que se extiende hasta 20 kHz, se cubre hasta 25 bandas críticas, cuyos anchos de banda individuales aumentan con la frecuencia. En términos generales, el sistema auditivo puede modelarse como un *banco de filtros pasobanda,* compuesto por 25 filtros pasobanda que se traslapan y que tienen anchos de banda menores que 100 Hz para las frecuencias audibles más bajas y hasta de 5 kHz para las más altas.

2. *Enmascaramiento auditivo.*
El enmascaramiento auditivo o enmascaramiento de ruido es un fenómeno en el dominio de la frecuencia que surge cuando una señal de bajo nivel (el enmascarador) y una señal de nivel alto (el enmascarado) ocurren simultáneamente y están lo suficientemente próximas en frecuencia. Si la señal de bajo nivel se encuentra por debajo del *umbral de enmascaramiento,* se hace inaudible (es decir, se enmascara) por efecto de la señal más intensa. El fenómeno de enmascaramiento auditivo es más notable cuando ambas señales se encuentran en la misma banda crítica, y menos efectivo cuando están en bandas vecinas.

La figura 3.33 ilustra la definición del umbral de enmascaramiento y los parámetros relacionados para un par de frecuencias adyacentes. Se supone que el enmascarador (es decir, la señal de alto nivel) se encuentra dentro de la banda crítica de sombreado oscuro. Las señales de bajo nivel se ubican dentro de esta área oscura y por debajo del umbral de enmascaramiento resultan enmascaradas por la señal más intensa. En la figura 3.33 vemos que el umbral de enmascaramiento varía con la frecuencia a lo largo de la banda crítica. Por tanto, es posible definir un *umbral de enmascaramiento crítico* para una banda crítica, debajo de la cual todas las señales de alto nivel que se encuentran dentro de la banda se hacen inaudibles por efecto de la señal más intensa. La diferencia de potencia, expresada en decibeles, entre el enmascarado y el umbral de enmascaramiento mínimo, se conoce como la *relación de señal a enmascaramiento* (SMR). La figura 3.33 incluye también la *relación señal a ruido* (SNR) para un cuantizador de R bits. La diferencia entre la SMR y la SNR es la *relación ruido a enmascaramiento* (*NMR*) para un cuantizador de R bits, como se indica mediante

$$\text{NMR} = \text{SMR} - \text{SNR} \tag{3.85}$$

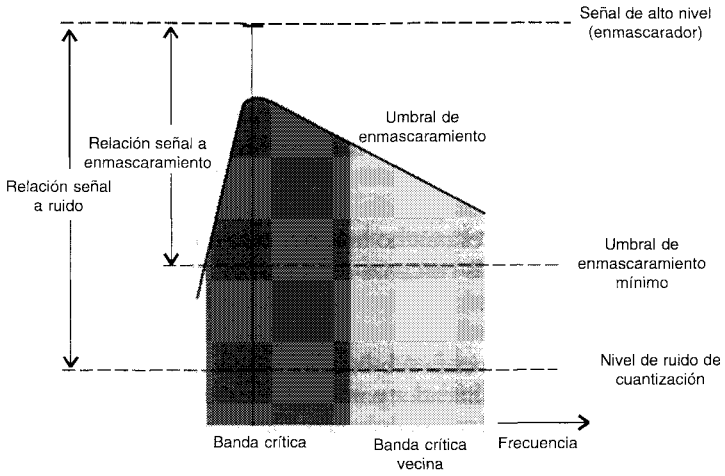


FIGURA 3.33 Ilustración de las definiciones del umbral de enmascaramiento y parámetros relacionados. La señal de alto nivel (enmascarador) se encuentra dentro de la banda crítica de sombreado oscuro, de ahí que el enmascaramiento resulta más efectivo en esta banda que en la banda vecina que se indica con un sombreado menor. (Adaptada de Noll, con autorización de la CRC Press.)

donde los tres términos se expresan en dB. Dentro de una banda crítica, el ruido de cuantización es inaudible siempre y cuando sea negativa la relación de ruido a enmascaramiento para el cuantizador pertinente.

Con estos fundamentos acerca de la psicoacústica del sistema auditivo, estamos listos para describir la operación del estándar de codificación de audio MPEG-1. La figura 3.34 muestra los diagramas de bloque básicos del codificador y el decodificador. El codificador consiste en cuatro unidades funcionales: red de correspondencia tiempo-frecuencia, modelo psicoacústico, cuantizador y codificador, y unidad de empaquetamiento de tramas. El decodificador está compuesto por tres unidades funcionales: unidad de desempaquetamiento de tramas, red de reconstrucción de frecuencia-muestra y red de correspondencia de frecuencia-tiempo. De esa forma, el modelo psicoacústico sólo es necesario en el codificador.

Empezando primero con la descripción del codificador, la función de la *red de correspondencia tiempo-frecuencia* es descomponer la señal de audio de entrada en sub-bandas múltiples para la codificación. La correspondencia se realiza en tres capas denominadas I, II y III, las

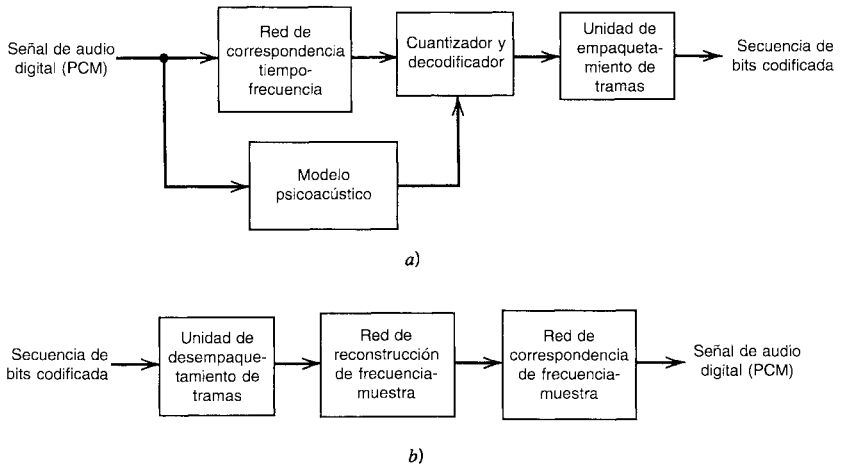


FIGURA 3.34 Sistema de codificación de audio MPEG. *a*) Transmisor. *b*) Receptor.

cuales son de complejidad, retardo y desempeño perceptual subjetivo crecientes. El algoritmo en la capa I utiliza un banco de filtros pasobanda que divide la señal de audio en 32 *sub-bandas de ancho constante*; este banco de filtros se encuentra también en las capas II y III. A la luz de nuestras notas anteriores acerca de las bandas críticas espaciadas de manera no uniforme, el diseño de este banco de filtros constituye un compromiso entre la eficiencia de cómputo y el desempeño perceptual. El algoritmo en la capa II es un simple mejoramiento de la capa I; mejora el desempeño de la compresión codificando los datos en grupos más grandes. Por último, el algoritmo de la capa III es mucho más refinado en cuanto a que se diseña para conseguir resoluciones de frecuencia más cercanas a las particiones entre las bandas críticas.

El *modelo psicoacústico* es el componente clave en el codificador. Su función estriba en analizar el contenido espectral de la señal de audio de entrada y, por ello, en calcular la relación señal a enmascaramiento para cada sub-banda en cada una de las tres capas. Esta información, a su vez, la usa el cuantizador-codificador para decidir cómo distribuir el número de bits disponible para la cuantización de las señales de sub-banda. Dicha asignación dinámica de bits se realiza de tal modo que se minimice la audición del ruido de cuantización. Por último, la *unidad de empaquetamiento de tramas* ensambla las muestras de audio cuantizadas en una secuencia de bits decodificable.

El decodificador recibe simplemente las operaciones de procesamiento de señales que se realizan en el codificador, convirtiendo la secuencia recibida de bits codificados en una señal de audio en el dominio del tiempo.

En resumen, el estándar de codificación de audio MPEG-1 representa la vanguardia en la codificación de señales de audio. La capa I alcanza una relación de compresión de 4 con respecto a una tasa de bits estereofónica aproximada de 384 kb/s para una calidad transparente del desempeño. Las relaciones de compresión correspondientes para las capas II y III son, en forma respectiva, 8 y 12 a tasas de bits estereofónicas aproximadas de 192 kb/s y 128 kb/s. La calidad subjetiva del estándar de codificación de audio MPEG-1 es equivalente a la calidad de los discos compactos (PCM de 16 bits) para muchos tipos de música; en la actualidad, el disco compacto (CD) es de hecho el estándar de la representación del audio digital.

3.18 *Resumen y discusión*

En este capítulo se presentan dos procesos fundamentales y complementarios:

- El *muestreo*, que opera en el dominio del tiempo; el proceso de muestreo constituye el enlace entre una forma de onda analógica y su representación en forma discreta.
- La *cuantización*, que opera en el dominio de la amplitud; el proceso de cuantización es el enlace entre una forma de onda analógica y su representación en amplitud discreta.

El proceso de muestreo se fundamenta en el *teorema de muestreo*, que establece que una señal estrictamente limitada en banda sin componentes de frecuencia superiores que W Hz se representa en forma única por una secuencia de muestras tomadas a una frecuencia uniforme igual o mayor que la frecuencia de Nyquist de $2W$ muestras por segundo. El proceso de cuantización aprovecha el hecho de que cualquier sentido humano, como receptor final, sólo puede detectar diferencias de intensidad finitas.

El proceso de muestreo resulta básico para la operación de todos los sistemas de modulación por pulsos, los cuales pueden clasificarse en modulación analógica por pulsos y modulación digital por pulsos. El rasgo distintivo entre ellos es que los sistemas de modulación analógica por pulsos mantienen una representación de amplitud continua de la señal de mensaje, en tanto que los sistemas de modulación digital por pulsos emplean también la cuantización para proporcionar una representación de la señal de mensaje que es discreta tanto en tiempo como en amplitud.

La modulación analógica por pulsos se produce al variar algún parámetro de los pulsos transmitidos, como la amplitud, duración o posición, en cuyo caso hablamos de modulación por amplitud de pulso (PAM), modulación por duración de pulso (PDM) o modulación por posición de pulso (PPM), respectivamente. En el multiplexado por división del tiempo (TDM) de

varios canales, el procesamiento de las señales suele iniciarse con la PAM. Para utilizar la PDM o la PPM en una aplicación de este tipo, tenemos que asegurar que la modulación a escala completa no ocasione que un pulso de una señal de mensaje entre a la franja del tiempo que pertenece a otra señal de mensaje. Esta restricción origina un uso ineficaz del espacio de tiempo de los sistemas telefónicos que se caracterizan por elevados factores pico, lo cual es una de las razones por las que no se usan la PDM o la PPM en la telefonía. Además, a pesar de que la PPM es más eficiente que la PDM, ambas están muy lejos del sistema ideal para intercambiar ancho de banda de transmisión por desempeño mejorado ante el ruido.

Los sistemas de modulación digital por pulsos transmiten señales de mensaje analógicas como una secuencia de pulsos codificados, lo cual se hace posible a través del uso combinado de muestreo y la cuantización. La modulación por codificación de pulsos es una forma importante de modulación digital por pulsos a la que se dota con algunas ventajas de sistema únicas, las cuales, simultáneamente, la han convertido en un método estándar de modulación para la transmisión de señales analógicas, como las de voz y video. Las ventajas de la modulación por codificación de pulsos incluyen robustez al ruido y la interferencia, regeneración eficiente de los pulsos codificados a lo largo de la trayectoria de transmisión y un formato uniforme para diferentes tipos de señales de banda base.

En realidad, las únicas ventajas de la modulación por codificación de pulsos se deben a que esta lista se ha vuelto el método preferido para la construcción de redes conmutadas de telefonía pública (PSTN). En este contexto, el lector debe advertir con todo cuidado que el canal telefónico visto a partir de la PSTN como un proveedor de servicios de Internet, por ejemplo, es *no lineal* debido al uso de la compresión y la expansión (compander), y, más importante, es *enteramente digital.* Esta observación tiene un impacto importante en el diseño de módems de alta velocidad para la comunicación entre un usuario de computadora y un servidor, como se explica en el capítulo 6.

La modulación delta y la modulación diferencial por codificación de pulsos son otras dos formas útiles de modulación digital por pulsos. La ventaja principal de la modulación delta es la simplicidad de su circuitería. En contraste, la modulación diferencial por codificación de pulsos emplea una complejidad de circuitería mayor para reducir el ancho de banda del canal. La mejora se logra utilizando la idea de predicción para eliminar los símbolos redundantes de una secuencia de datos entrante. Es posible una mejora adicional en la operación de la modulación diferencial por codificación de pulsos a partir del uso de la adaptabilidad para tener en cuenta las variaciones estadísticas de los datos de entrada. Al hacerlo así, el requerimiento de ancho de banda se reduce de manera considerable sin una degradación seria del desempeño del sistema.

En contraste con la modulación diferencial adaptable por codificación de pulsos, el estándar de codificación de audio MPEG consigue la compresión de señales de audio estereofónicas de manera transparente y sin pérdidas perceptuales. Este impresionante desempeño se realiza explotando ciertas propiedades psicoacústicas del sistema auditivo.

En este punto de la exposición resulta informativo llevar a cabo un análisis crítico de las diferentes formas de la modulación por pulsos que se han descrito en este capítulo. En sentido estricto, el término *modulación por pulsos* es un nombre inapropiado porque todas sus formas diferentes, ya sean analógicas o digitales, son de hecho técnicas de *codificación de fuente.* Afirmamos lo anterior por la simple razón de que una señal de mensaje sigue siendo una señal de banda base después de experimentar todos los cambios implicados en un proceso de modulación por pulsos. La naturaleza de banda base de una señal modulada por pulsos se ejemplifica porque, sin que importe su descripción exacta, es posible transmitirla por un canal de banda base de ancho de banda adecuado. En realidad, el material que se presenta en el siguiente capítulo se dedica a la transmisión de banda base de datos representados por una secuencia de pulsos.

Es importante admitir que las técnicas de modulación por pulsos presentan *pérdidas* a causa de que algo de la información se pierde como consecuencia de la representación de señales que realizan. Por ejemplo, en la modulación por amplitud de pulso, la práctica usual implica el uso del filtrado antitraslape pasobajas antes del muestreo. De ese modo, la información se pierde porque las compo-

nentes de alta frecuencia consideradas como esenciales son eliminadas por el filtro. La naturaleza con pérdida de la modulación por pulsos se advierte en forma más tangible en la modulación por codificación de pulsos que se caracterizan por la generación de ruido de cuantización (es decir, distorsión). La secuencia transmitida de pulsos codificados no tiene la precisión infinita que se necesita para representar exactamente muestreos continuos. A pesar de todo, la pérdida de información en que se incurre en un proceso de modulación por pulsos está *bajo el control del diseñador* en cuanto a que es posible hacerla lo suficientemente pequeña e imperceptible para el usuario final.

El material que se presenta en este capítulo acerca de la modulación por pulsos ha surgido de la perspectiva del procesamiento de señales. Volveremos a la modulación por codificación de pulsos en el capítulo 9, el cual se dedica a las consideraciones teóricas de la información de los sistemas de comunicación. De esa manera estableceremos un conocimiento más profundo de su operación como una técnica de codificación de fuente.

NOTAS Y BIBLIOGRAFÍA

1. El libro clásico acerca de la modulación por pulsos es el de Black [22]. Un tratamiento más detallado del tema se presenta en el libro de Rowe [130]. En relación con el análisis de ruido de un sistema PPM, vea la tercera edición del libro de Haykin [61].
2. La modulación por codificación de pulsos la inventó Reeves en 1937. En relación con una explicación histórica de esta invención, vea el artículo de Reeves [350]. El libro de Jayant y Noll [71] presenta un tratamiento detallado de la modulación por codificación de pulsos, la modulación diferencial por codificación de pulsos, la modulación delta y sus variantes. El libro editado por Jayant [72] proporciona una colección de artículos anteriores escritos sobre la cuantización y codificación de formas de onda.
3. Para una explicación detallada del ruido de cuantización en sistemas PCM, vea el artículo de Bennett [201] y el libro de Rowe ([130], pp. 311-321).
4. Las dos condiciones necesarias de las ecuaciones (3.42) y (3.47) con respecto a la optimización de un cuantizador escalar fueron informadas de manera independiente por Lloyd [310] y Max [319], y de ahí el nombre "cuantizador Lloyd-Max". La deducción de estas dos condiciones de optimización presentada en este capítulo se obtuvo con base en el libro de Gersho y Gray [50].
5. La ley μ utilizada para la compresión de señales se describe en Smith [376]. La ley μ se utiliza en Estados Unidos, Canadá y Japón. En Europa se emplea la ley A para la compresión de señales; esta ley de compresión se describe en Cattermole ([28], pp. 133-140). Con respecto a una explicación de la ley μ y de la ley A, vea también el artículo de Kaneko [294].
6. Para una descripción de la versión original del sistema de portadora T1, vea el artículo de Fulz y Penick [253]. La descripción que se da en el ejemplo 3.2 se basó en una versión actualizada de este sistema; consulte Hanning y Pan [276].
7. Los estándares estadounidenses/japoneses para multiplexores digitales fueron adoptados originalmente por AT&T. La CCITT ha adoptado otro conjunto de estándares para el resto del mundo. La jerarquía digital CCITT es similar a la que se describe en la sección 3.10, salvo por ciertos cambios en las especificaciones del número de entrada del canal a los cinco multiplexores digitales y sus tasas de bits individuales. En relación con detalles de la jerarquía digital CCITT, vea Couch [32].
8. En lo que respecta a los artículos originales acerca de la modulación delta, vea Schouten, DeJager y Greefkes [368] y DeJager [230]. Con respecto a un artículo retrospectivo acerca de la modulación delta, consulte el escrito de Schindler [365].
9. La modulación delta-sigma se describe en el libro de Jayant y Noll ([71], pp. 399-400); vea también el artículo de Insoe, Yasuda y Murakami [283].
10. Cutler inventó la modulación diferencial por codificación de pulsos; la misma se describe en una patente publicada en 1952. Para una comparación de los desempeños con respecto al ruido de la PCM y la DPCM, vea el artículo de Jayant [287]; consulte también Rabiner y Schafer ([124], capítulo 5).

11. En relación con una discusión de la modulación diferencial adaptable por codificación de pulsos, vea Jayant y Noll [71].
12. El algoritmo de modulación delta adaptable (ADM) de la ecuación (3.84) es la versión corregida de un algoritmo presentado en Sklar ([140], p. 641). El algoritmo de Sklar se adaptó con base en un artículo anterior de Song y otros, donde se obtiene un sistema ADM óptimo; las ecuaciones altamente no lineales que caracterizan al sistema óptimo se aproximan en un artículo posterior mediante ecuaciones lineales por secciones con fines de implantación.
13. El estándar de codificación de audio MPEG-1 se describe en los artículos de Brandenburg y Stoll [208], Pan [336] y el de Peter Noll en el manual Digital Signal Processing editado por Madisetti y Williams [99]; el último artículo analiza también los estándares que siguieron al MPEG-1. En particular, el MPEG-2 ofrece codificación de audio estereofónica a frecuencias de muestreo inferiores a las del MPEG-1.
14. El oído, órgano de la audición, responde a ondas acústicas entrantes. Tiene tres partes principales cuyas funciones se resumen a continuación:
    - El *oído externo* ayuda en la colección de sonidos.
    - El *oído medio* proporciona un acoplamiento de impedancia acústica entre el aire y los fluidos de la cóclea, conduciendo de ese modo y de manera eficiente las vibraciones de la membrana timpánica (tímpano) debidas a los sonidos entrantes al oído interno.
    - El *oído interno* convierte las vibraciones mecánicas provenientes del oído medio en una señal electroquímica o nerviosa para la transmisión al cerebro y el procesamiento.

PROBLEMAS

Proceso de muestreo

3.1 Una señal de banda angosta tiene un ancho de banda de 10 kHz centrado en una frecuencia de portadora de 100 kHz. Se propone representar esta señal en una forma en tiempo discreto mediante el muestreo individual de sus componentes en fase y en cuadratura. ¿Cuál es la frecuencia de muestreo mínima que puede utilizarse en esta representación? Justifique su respuesta. ¿Cómo reconstruiría la señal de banda angosta individual a partir de las versiones muestreadas de sus componentes en fase y en cuadratura?

3.2 En el *muestreo natural* se multiplica una señal analógica $g(t)$ por un tren periódico de pulsos rectangulares $c(t)$. Debido a que la frecuencia de repetición de pulsos de este tren periódico es f_s y que la duración de cada pulso rectangular es T (con $f_s T \ll 1$), realice lo siguiente:

a) Encuentre el espectro de la señal $s(t)$ que resulta del uso del muestreo natural; puede suponer que el tiempo $t = 0$ corresponde al punto medio de un pulso rectangular en $c(t)$.

b) Muestre que la señal original $m(t)$ puede recuperarse exactamente de su versión muestreada en forma natural siempre y cuando se satisfagan las condiciones incorporadas en el teorema de muestreo.

3.3 Especifique la frecuencia de Nyquist y el intervalo de Nyquist para cada una de las siguientes señales:

a) $g(t) = \text{senc}(200t)$

b) $g(t) = \text{senc}^2(200t)$

c) $g(t) = \text{senc}(200t) + \text{senc}^2(200t)$

3.4 **a)** Grafique el espectro de una onda PAM producida por la señal moduladora

$$m(t) = A_m \cos(2\pi f_m t)$$

suponiendo una frecuencia de modulación $f_m = 0.25$ Hz, periodo de muestreo $T_s = 1$ s y duración de pulso $T = 0.45$ s.

b) Utilizando un filtro de reconstrucción ideal, grafique el espectro de la salida del filtro. Compare este resultado con la salida que se obtendría si no hubiera efecto de apertura.

Modulación por amplitud de pulso

3.5 La figura P3.5 muestra el espectro idealizado de una señal de mensaje $m(t)$. La señal se muestrea a una frecuencia igual a 1 kHz utilizando pulsos de cresta plana, siendo cada pulso de amplitud unitaria y duración igual a 0.1 ms. Determine y dibuje el espectro de la señal PAM resultante.

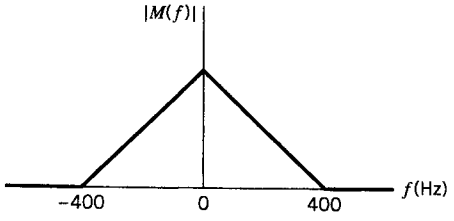


FIGURA P3.5

3.6 En este problema evaluamos la ecualización que se necesita para el efecto de apertura en un sistema PAM. La frecuencia de operación es $f = f_s/2$, la cual corresponde al componente de frecuencia más alto de la señal de mensaje para una frecuencia de muestreo igual a la de Nyquist. Grafique $1/\text{senc}(0.5T/T_s)$ en función de T/T_s y, en consecuencia, determine la ecualización necesaria cuando $T/T_s = 0.1$.

3.7 Considere una onda PAM transmitida por un canal con ruido blanco gaussiano y con ancho de banda mínimo $B_T = 1/2T_s$, donde T_s es el periodo de muestreo. El ruido es de media cero y densidad de potencia espectral $N_0/2$. La señal PAM utiliza un pulso estándar $g(t)$ con su transformada de Fourier definida por

$$G(f) = \begin{cases} \dfrac{1}{2B_T}, & |f| < B_T \\ 0, & |f| > B_T \end{cases}$$

Considerando una onda moduladora senoidal a carga máxima, demuestre que la PAM y la transmisión de señal de banda base tienen iguales relaciones señal a ruido para la misma potencia transmitida promedio.

3.8 Se muestrean 24 señales de voz de manera uniforme y luego se multiplexan por división de tiempo. La operación de muestreo utiliza muestras de cresta plana con duración de 1 μs. La operación de multiplexado incluye la provisión de sincronización para agregar un pulso extra de amplitud suficiente y también de 1 μs de duración. La componente de más alta frecuencia de cada señal de voz es de 3.4 kHz.

a) Suponiendo una frecuencia de muestreo de 8 kHz, calcule el espaciamiento entre pulsos sucesivos de la señal multiplexada.

b) Repita su cálculo asumiendo el empleo de muestreo a la frecuencia de Nyquist.

3.9 Doce señales de mensaje diferentes, cada una con un ancho de banda de 10 kHz, se van a multiplexar y transmitir. Determine el ancho de banda mínimo que se requiere en cada caso si el método de multiplexado/modulación que se usa es

a) FDM, SSB.

b) TDM, PAM.

3.10 Un sistema de telemetría PAM incluye el multiplexado de cuatro señales de entrada: $s_i(t)$, $i = 1, 2, 3, 4$. Dos de las señales, $s_1(t)$ y $s_2(t)$, tienen anchos de banda cada una de 80 Hz, en tanto que las dos señales restantes, $s_3(t)$ y $s_4(t)$, tienen cada una anchos de banda de 1 kHz. Cada una de las señales $s_3(t)$ y $s_4(t)$ se muestrean a razón de 2400 muestras por segundo. Esta frecuencia de muestreo se divide entre 2^R (es decir, una potencia entera de 2) para obtener la frecuencia de muestreo de $s_1(t)$ y $s_2(t)$.

a) Determine el valor máximo de R.

b) Utilizando el valor de R que se encontró en la parte *a*, diseñe un sistema de multiplexado que multiplexa primero $s_1(t)$ y $s_2(t)$ en una nueva secuencia, $s_5(t)$, y que luego multiplexa a $s_3(t)$, $s_4(t)$ y $s_5(t)$.

Códigos de línea

3.11 En este problema obtenemos las fórmulas que se utilizan para calcular los espectros de potencia de la figura 3.16 correspondientes a los cinco códigos de línea que se describen en la sección 3.7. En el caso de cada código de línea, la duración del bit es T_b y la amplitud del pulso A se acondiciona para normalizar la potencia promedio del código de línea a la unidad, como se indica en la figura 3.16. Suponga que la secuencia de datos se genera aleatoriamente, y que los símbolos 0 y 1 son igualmente probables.

Obtenga las densidades espectrales de potencia de estos códigos de línea como se resume a continuación:

a) Señales unipolares sin retorno a cero:

$$S(f) = \frac{A^2 T_b}{4} \text{senc}^2(fT_b)\left(1 + \frac{1}{T_b}\delta(f)\right)$$

b) Señales polares sin retorno a cero:

$$S(f) = A^2 T_b \text{ senc}^2(fT_b)$$

c) Señales unipolares sin retorno a cero:

$$S(f) = \frac{A^2 T_b}{16} \text{senc}^2\left(\frac{fT_b}{2}\right)\left[1 + \frac{1}{T_b}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right]$$

d) Señales bipolares con retorno a cero:

$$S(f) = \frac{A^2 T_b}{4} \text{senc}^2\left(\frac{fT_b}{2}\right)\text{sen}^2(\pi f T_b)$$

e) Señales con codificación de Manchester

$$S(f) = A^2 T_b \text{senc}^2\left(\frac{fT_b}{2}\right)\text{sen}^2\left(\frac{\pi f T_b}{2}\right)$$

Por tanto, confirme las gráficas espectrales que se ilustran en la figura 3.16.

3.12 Suponga que una secuencia aleatoria de datos binarios (con símbolos igualmente probables) se codifica diferencialmente y se transmite después utilizando uno de los cinco códigos de línea descritos en el problema 3.11. ¿Cómo afecta el uso de la codificación diferencial a la densidad espectral de potencia de los datos transmitidos? Justifique su respuesta.

3.13 Una secuencia de datos generados al azar consiste en símbolos 0 y 1 binarios igualmente probables y se codifica en una forma de onda polar sin retorno a cero, definiéndose cada símbolo binario del modo siguiente:

$$s(t) = \begin{cases} \cos\left(\dfrac{\pi t}{T_b}\right), & -\dfrac{T_b}{2} < t \leq \dfrac{T_b}{2} \\ 0, & \text{en otro caso} \end{cases}$$

a) Dibuje la forma de onda generada de ese modo, suponiendo que la secuencia de datos es 00101110.

b) Obtenga una expresión para la densidad espectral de potencia de esta señal y dibújela. *Sugerencia:* emplee la ecuación (1.52).

c) Compare la densidad espectral de potencia de esta forma de onda aleatoria con la que se definió en la parte *b* del problema 3.11.

3.14 Dada la secuencia de datos 1110010100, dibuje la secuencia de pulsos transmitida para cada uno de los siguientes códigos de línea:

a) Unipolares sin retorno a cero

b) Polares sin retorno a cero

c) Unipolares con retorno a cero

d) Bipolares con retorno a cero

e) Código Manchester

3.15 Suponga que la secuencia de datos binarios que se considera en el problema 3.14 se codifica diferencialmente y luego se transmite empleando uno de los cinco códigos de línea considerados ahí. Dibuje cada una de las secuencias de datos transmitidos, suponiendo el símbolo 1 como el bit de referencia. ¿Cómo se afecta el resultado si el símbolo 0 se usa como el bit de referencia?

Modulación por codificación de pulsos

3.16 Una señal de voz tiene una duración total de 10 s. Se muestrea a una frecuencia de 8 kHz y luego se codifica. Se requiere que la relación señal a ruido (cuantización) sea de 40 dB. Calcule la capacidad de almacenamiento mínimo que se necesita para alojar esta señal de voz digitalizada.

3.17 Considere un cuantizador uniforme caracterizado por la relación de entrada-salida que se ilustra en la figura 3.10*a*. Suponga que una variable aleatoria distribuida gaussianamente con media cero y varianza unitaria se aplica a esta entrada del cuantizador.

a) ¿Cuál es la probabilidad de que la amplitud de la entrada se encuentre fuera del intervalo de -4 a $+4$?

b) Utilizando el resultado de la parte *a*, demuestre que la relación señal a ruido de salida del cuantizador está dada por

$$(\text{SNR})_O = 6R - 7.2 \text{ dB}$$

donde R es el número de bits por muestra. De manera específica, puede suponer que la entrada del cuantizador se extiende de -4 a $+4$. Compare el resultado de la parte *b* con el que se obtuvo en el ejemplo 3.1.

3.18 Un sistema PCM utiliza un cuantizador uniforme seguido por un codificador binario de 7 bits. La tasa de bits del sistema es igual a 50×10^6 b/s.

a) ¿Cuál es el máximo ancho de banda del mensaje para el cual el sistema opera en forma satisfactoria?

b) Determine la relación señal a ruido de salida (cuantización) cuando se aplica a la entrada una onda moduladora senoidal a máxima carga de frecuencia igual a 1 MHz.

3.19 Demuestre que con un cuantizador no uniforme el valor cuadrático medio del error de cuantización es aproximadamente igual a $(1/12)\Sigma_i \Delta_i^2 p_i$, donde Δ_i es el tamaño de escalón i-ésimo y p_i es la probabilidad de que la señal de entrada se encuentre dentro del intervalo i-ésimo. Suponga que el tamaño del escalón Δ_i es pequeño en comparación con la excursión de la señal de entrada.

3.20 **a)** Una señal senoidal, con una amplitud de 3.25 volts, se aplica a un cuantizador uniforme del tipo de medio escalón cuya salida toma los valores 0, ± 1, ± 2, ± 3 volts. Dibuje la forma de onda de la salida de cuantizador que resulta para un ciclo completo de la entrada.

b) Repita esta evaluación para el caso en el que el cuantizador es del tipo de medio aumento y cuya salida toma los valores $= 0.5, \pm 1.5, \pm 2.5, \pm 3.5$ volts.

3.21 La señal

$$m(t) = 6 \operatorname{sen}(2\pi t) \text{ volts}$$

se transmite utilizando un sistema PCM binario de 4 bits. El cuantizador es del tipo de medio aumento, con tamaño de escalón de 1 volt. Dibuje la onda PCM resultante para un ciclo completo de la entrada. Suponga una frecuencia de muestreo de 4 muestras por segundo, con muestras que se toman a $t = \pm 1/8, \pm 3/8, \pm 5/8, \ldots$, segundos.

3.22 La figura P3.22 muestra una señal PCM en la cual se usan niveles de amplitud de $+1$ volt y -1 volt para representar símbolos binarios 1 y 0, respectivamente. La palabra de código utilizada consta de tres bits. Determine la versión muestreada de una señal analógica a partir de la cual se obtiene esta señal PCM.

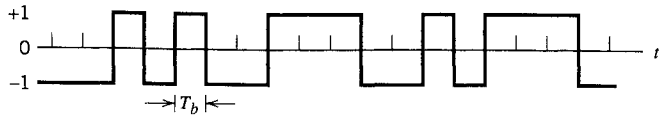


FIGURA P3.22

3.23 Considere una cadena de $(n-1)$ repetidores regenerativos, con un total de n decisiones secuenciales efectuadas sobre una onda PCM binaria, incluida la decisión final que se realiza en el receptor. Suponga que cualquier símbolo binario transmitido por el sistema tiene una probabilidad independiente p_1 de ser invertido por cualquier repetidor. Deje que p_n represente la probabilidad de que un símbolo binario sea erróneo después de la transmisión a través del sistema completo.

a) Demuestre que

$$p_n = \frac{1}{2}\left[1-(1-2p_1)^n\right]$$

b) Si p_1 es muy pequeño y n no es muy grande, ¿cuál es el valor correspondiente de p_n?

3.24 Analice los aspectos básicos implicados en el diseño de un repetidor regenerativo para la modulación por codificación de pulsos.

Modulación delta

3.25 Considere una señal de prueba $m(t)$ definida por una función tangente hiperbólica:

$$m(t) = A\tanh(\beta t)$$

donde A y β son constantes. Determine el tamaño de escalón mínimo Δ para la modulación delta de esta señal, la cual se requiere para evitar la sobrecarga de pendiente.

3.26 Considere una onda seno de frecuencia f_m y amplitud A_m, que se aplica a un modulador delta de tamaño de escalón Δ. Demuestre que ocurrirá la distorsión de sobrecarga de pendiente si

$$A_m > \frac{\Delta}{2\pi f_m T_s}$$

donde T_s es el periodo de muestreo. ¿Cuál es la potencia máxima que es posible transmitir sin distorsión de sobrecarga de pendiente?

3.27 Se diseña un modulador delta lineal para operar sobre señales de voz limitadas a 3.4 kHz. Las especificaciones del modulador son las siguientes:

- Frecuencia de muestreo = $10f_{\text{Nyquist}}$, donde f_{Nyquist} es la frecuencia de Nyquist de la señal de voz.
- Tamaño del escalón Δ = 100 mV.

El modulador se prueba con una señal senoidal de 1 kHz. Determine la amplitud máxima de esta señal de prueba requerida para evitar la sobrecarga de pendiente.

3.28 En este problema obtenemos una forma empírica para la relación promedio señal a ruido (cuantización) de un sistema DM con una señal senoidal de amplitud A y frecuencia f_m como la señal de prueba. Suponga que la densidad espectral de potencia del ruido granular generado por el sistema está gobernada por la fórmula

$$S_N(f) = \frac{\Delta^2}{6f_s}$$

donde f_s es la frecuencia de muestreo y Δ es el tamaño de escalón. (Advierta que esta fórmula es básicamente la misma que la de la densidad espectral de potencia del ruido de cuantización en un

sistema PCM con la $\Delta/2$ correspondiente a PCM sustituyéndose por la Δ relativa a DM.) El sistema DM se diseña para manejar señales de mensaje analógicas limitadas al ancho de banda W.

a) Demuestre que la potencia promedio del ruido de cuantización producida por este sistema es

$$N = \frac{4\pi^2 A^2 f_m^2 W}{3f_s^3}$$

donde se supone que el tamaño de escalón Δ sea elegido de acuerdo con la fórmula utilizada en el problema 3.27 a fin de evitar la sobrecarga de pendiente.

b) Por tanto, determine la relación señal a ruido (cuantización) del sistema DM para una entrada senoidal.

3.29 Considere un sistema DM diseñado para alojar señales de mensaje analógicas limitadas a un ancho de banda $W = 5$ kHz. Una señal de prueba senoidal de amplitud $A = 1$ volt y frecuencia $f_m = 1$ kHz se aplica al sistema. La frecuencia de muestreo del sistema corresponde a 50 kHz.

a) Calcule el tamaño de escalón Δ requerido para minimizar la sobrecarga de pendiente.

b) Calcule la relación señal a ruido (cuantización) del sistema para la señal de prueba senoidal especificada.

En estos cálculos utilice las fórmulas que se obtuvieron en los problemas 3.27 y 3.28.

3.30 Considere una señal pasobajas con un ancho de banda de 3 kHz. Se usa un sistema de modulación delta lineal, con tamaño de escalón $\Delta = 0.1$ V, para procesar esta señal a una frecuencia de muestreo igual a 10 veces la frecuencia de Nyquist.

a) Evalúe la amplitud máxima de la señal senoidal de prueba de frecuencia igual a 1 kHz, que el sistema puede procesar sin distorsión de sobrecarga de pendiente.

b) Para la especificación indicada en la parte *a*, evalúe la relación señal a ruido de salida en condiciones 1) de prefiltrado y 2) de posfiltrado.

Predicción lineal

3.31 Un predictor lineal de un escalón opera sobre una versión muestreada de una señal senoidal. La frecuencia de muestreo es igual a $10f_0$, donde f_0 es la frecuencia de la senoide. El predictor tiene un solo coeficiente denotado por w_1.

a) Determine el valor óptimo de w_1 que se requiere para minimizar la varianza del error de predicción.

b) Determine el valor mínimo de la varianza del error de la predicción.

3.32 Un proceso estacionario $X(t)$ tiene los siguientes valores para su función de autocorrelación:

$$R_X(0) = 1$$
$$R_X(1) = 0.8$$
$$R_X(2) = 0.6$$
$$R_X(3) = 0.4$$

a) Calcule los coeficientes de un predictor lineal óptimo que incluya el uso de tres retardos unitarios.

b) Calcule la varianza del error que resulta de la predicción.

3.33 Repita los cálculos del problema 3.32, pero esta vez utilice un predictor lineal con dos retardos unitarios. Compare el desempeño de este segundo predictor lineal óptimo con el que se consideró en el problema 3.32.

Modulación diferencial por codificación de pulsos

3.34 Un sistema DPCM utiliza un predictor lineal con una sola derivación. La función de autocorrelación normalizada de la señal de entrada para un retardo de un intervalo de muestreo es de 0.75. El

predictor se diseña para minimizar la varianza del error de predicción. Determine la ganancia de procesamiento que se logra mediante el uso de este predictor.

3.35 Calcule el mejoramiento en la ganancia de procesamiento de un sistema DPCM utilizando el predictor lineal optimizado de tres derivaciones del problema 3.32 con respecto al predictor lineal optimizado de dos derivaciones del problema 3.33. Utilice en el cálculo los valores de la función de autocorrelación de la señal de entrada que se especifican en el problema 3.32.

3.36 En este problema comparamos el desempeño de un sistema DPCM con el de un sistema PCM ordinario utilizando la compresión-expansión.

Para un número suficientemente grande de niveles de representación, la relación señal a ruido (cuantización) de los sistemas PCM está definida, en general, por

$$10 \log_{10} (\text{SNR})_O = \alpha + 6n \text{ dB}$$

donde 2^n es el número de niveles de representación. Para un sistema PCM de compresión-expansión utilizando la ley μ, la propia constante α está definida por

$$\alpha \simeq 4.77 - 20 \log_{10} (\log(1+\mu)) \text{ dB}$$

En un sistema DPCM, en cambio, la constante α se encuentra en el intervalo $-3 < \alpha < 15$ dB. Las fórmulas anotadas aquí se aplican a señales del habla de calidad telefónica.

Compare el desempeño de un sistema DPCM contra el de un sistema PCM comprimido-expandido μ, con $\mu = 255$ en cada uno de los siguientes escenarios:

a) La mejora en $(\text{SNR})_O$ realizada por la DPCM sobre la PCM de compresión-expansión para el mismo número de bits por muestra.

b) La reducción en el número de bits por muestra que requiere la DPCM, en comparación con la PCM de compresión-expansión para la misma $(\text{SNR})_O$.

3.37 En el sistema DPCM que se describe en la figura P3.37, demuestre que a falta de ruido del canal, los filtros de predicción de transmisión y recepción operan sobre señales de entrada ligeramente diferentes.

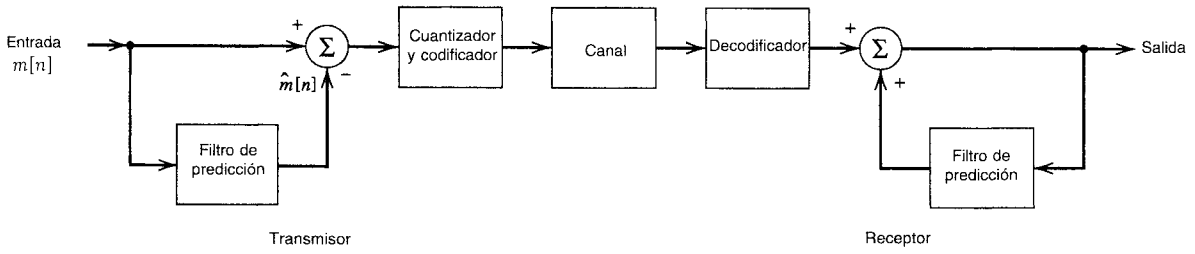


FIGURA P3.37

Experimentos de computadora

3.38 Una señal senoidal de frecuencia $f_0 = 10^4/2\pi$ Hz se muestrea a la velocidad de 8 kHz y después se aplica a un circuito de muestreo y retención para producir una señal PAM de cresta plana $s(t)$ con duración de pulso $T = 500$ μs.

a) Calcule la forma de onda de la señal PAM $s(t)$.

b) Calcule $|S(f)|$, que denota el espectro de magnitud de la señal PAM $s(t)$.

c) Calcule la envolvente de $|S(f)|$. Por tanto, confirme que la frecuencia a la cual esta envolvente pasa por cero por primera vez es igual a $(1/T) = 20$ kHz.

3.39 En este problema utilizamos una simulación de computadora para comparar el desempeño de un sistema PCM de compresión-expansión usando la ley μ contra el del sistema correspondiente em-

pleando un cuantizador uniforme. La simulación se va a realizar en el caso de una señal de entrada senoidal de amplitud variable.

a) Empleando la ley μ que se describe en la tabla 3.4, grafique la relación señal a ruido de salida como una función de la relación señal a ruido de entrada, expresando ambas relaciones en decibeles.

b) Compare los resultados de su cálculo en la parte *a* con un cuantizador uniforme que tenga 256 niveles de representación.

3.40 En este experimento estudiamos la proyección adaptable lineal de una señal $x[n]$ gobernada por la siguiente repetición:

$$x[n] = 0.8x[n-1] - 0.1x[n-2] + 0.1v[n]$$

donde $v[n]$ se extrae de un proceso en tiempo discreto de ruido blanco, media cero y varianza unitaria. (Un proceso generado de esta manera se conoce como *proceso autorregresivo de orden dos.*) Específicamente, la predicción adaptable se lleva a cabo utilizando el *algoritmo de mínimos cuadrados normalizados* definido por

$$\hat{x}[n] = \sum_{k=1}^{p} w_k[n]x[n-k]$$

$$e[n] = x[n] - \hat{x}[n]$$

$$w_k[n+1] = w_k[n] + \left(\frac{\mu}{\sum_{k=1}^{p} x^2[n-k]} \right) x[n-k]e[n], \quad k = 1, 2, \ldots, p$$

donde p es el orden de predicción y μ es el parámetro normalizado de tamaño de escalón. El punto importante que hay que advertir aquí es que μ es adimensional y se asegura la estabilidad del algoritmo eligiéndolo de acuerdo con la fórmula

$$0 < \mu < 2$$

El algoritmo se inicia estableciendo

$$w_k[0] = 0 \qquad \text{para } k$$

La *curva de aprendizaje* del algoritmo se define como una gráfica del error cuadrático medio en función del número de iteraciones n para los valores especificados de los parámetros, que se obtienen promediando la gráfica de $e^2[n]$ en función de n a lo largo de un gran número de realizaciones diferentes del algoritmo.

a) Grafique las curvas de aprendizaje para la predicción adaptable de $x[n]$ correspondiente a un orden de predicción fijo $p = 5$ y tres valores diferentes del parámetro de tamaño de escalón: $\mu = 0.0075$, 0.05 y 0.5.

b) ¿Qué observaciones puede usted señalar a partir de las curvas de aprendizaje de la parte *a*?

CAPÍTULO 4

TRANSMISIÓN DE PULSOS EN BANDA BASE

En este capítulo explicamos la transmisión de datos digitales por un canal de la banda base, subrayando los siguientes temas:

- *El filtro de acoplamiento, que constituye el sistema óptimo para detectar una señal conocida en ruido blanco gaussiano aditivo.*
- *Cálculo de la tasa de error de bits por efecto de la presencia de ruido del canal.*
- *Interferencia de intersímbolo, la cual surge cuando el canal es dispersor, como suele ocurrir en la práctica.*
- *Criterio de Nyquist para la transmisión de datos por la banda base sin distorsión.*
- *Codificación de nivel correlativa o transmisión de señales en respuesta parcial para combatir los efectos de la interferencia de intersímbolo.*
- *Líneas digitales de suscriptor.*
- *Ecualización de un canal disperso de la banda base.*
- *El patrón del ojo para exhibir los efectos combinados de la interferencia de intersímbolo y el ruido del canal en la transmisión de datos.*

4.1 *Introducción*

En el capítulo 3 describimos técnicas para convertir una señal analógica que contiene información a la forma digital. Hay otra manera en la cual es posible que se presenten los datos digitales en la práctica: los datos quizá representen la salida de una fuente de información que es de naturaleza inherentemente discreta (por ejemplo, una computadora digital). En este capítulo estudiaremos la transmisión de datos digitales (de cualquier origen) por un *canal de banda base*.[1] La transmisión de datos por un canal pasobanda que utiliza modulación se aborda en el capítulo 6.

Los datos digitales tienen un amplio espectro con un importante contenido de baja frecuencia. Por tanto, la transmisión por banda base de datos digitales requiere el empleo de un canal pasobajas con un ancho de banda suficientemente grande para alojar el contenido de frecuencia esencial de la secuencia de datos. Ahora bien, por lo común el canal es *disperso*, en el sentido de que su respuesta en frecuencia se desvía de la de un filtro pasobajas ideal. El resultado de la transmisión de datos por un canal de ese tipo es que cada pulso recibido resulta afectado un poco por los pulsos adyacentes, originando con ello una forma común de interferencia que se conoce como *de intersímbolo* (ISI). Esta interferencia es una fuente principal de errores de bit en la secuencia de datos reconstruida a la salida del receptor. Para corregirla tiene que ejercerse control sobre la forma del pulso en todo el sistema. Por ello, buena parte del material de este capítulo se dedica de una manera u otra a la *formación del pulso*.

Otra fuente de errores de bit en un sistema de transmisión de datos de banda base corresponde al siempre presente *ruido del canal*. Naturalmente, el ruido y la ISI surgen simultáneamente en el sistema. Sin embargo, para comprender cómo afectan el desempeño del mismo, los consideramos primero por separado; más adelante en el capítulo estudiamos sus efectos combinados.

Iniciamos el capítulo describiendo un resultado fundamental en la teoría de comunicaciones, que tiene que ver con la *detección* de una señal de pulsos de forma de onda conocida que está inmersa en ruido blanco aditivo. El dispositivo para la detección óptima de un pulso de este tipo implica el uso de un filtro lineal e invariante con el tiempo conocido como *filtro de acoplamiento*,[2] el cual recibe ese nombre debido a que su respuesta al impulso se acopla a la señal del pulso.

4.2 *Filtro de acoplamiento*

Un problema básico que surge a menudo en el estudio de los sistemas de comunicación es la *detección* de un pulso transmitido por un canal corrompido por el ruido de este mismo (es decir, ruido aditivo en el extremo frontal del receptor). Para los fines de la exposición que se presenta en este apartado, suponemos que la fuente principal de limitación del sistema es el ruido del canal.

Considere entonces el modelo de receptor que se presenta en la figura 4.1, que incluye un filtro lineal e invariante con el tiempo de respuesta al impulso $h(t)$. La entrada del filtro $x(t)$ consiste en una señal de pulso $g(t)$ corrompida por el ruido aditivo del canal $w(t)$, como se indica por medio de

$$x(t) = g(t) + w(t), \qquad 0 \le t \le T \tag{4.1}$$

donde T es un intervalo de observación arbitrario. La señal de pulso $g(t)$ puede representar un símbolo binario 1 o 0 en un sistema de comunicación digital. El término $w(t)$ es la función de muestreo de un proceso de ruido blanco de media cero y densidad espectral de potencia $N_0/2$. Se supone que el receptor tiene conocimiento de la forma de onda de la señal de pulso $g(t)$. La fuente de incertidumbre se encuentra en el ruido $w(t)$. La función del receptor es detectar la señal de pulso $g(t)$ de manera óptima, dada la señal recibida $x(t)$. Para satisfacer este requisito tenemos que optimizar el diseño del filtro de modo que se minimicen los efectos del ruido a su salida en cierto sentido estadístico, y para mejorar de esa manera la detección de la señal de pulso $g(t)$.

Puesto que el filtro es lineal, la salida $y(t)$ que se produce puede expresarse como

$$y(t) = g_o(t) + n(t) \tag{4.2}$$

donde $g_o(t)$ y $n(t)$ son generadas respectivamente por las componentes de la señal y del ruido de la entrada $x(t)$. Una manera sencilla de describir el requerimiento de que la componente de la señal de salida $g_o(t)$ es considerablemente mayor que la componente del ruido de salida $n(t)$ consiste en lograr que el filtro haga lo más grande posible la potencia instantánea en la señal de salida $g_o(t)$ medida en el tiempo $t = T$, en comparación con la potencia promedio del ruido de salida $n(t)$. Esto equivale a maximizar la *relación señal a ruido del pulso pico*, definida como

$$\eta = \frac{|g_o(T)|^2}{E[n^2(t)]} \tag{4.3}$$

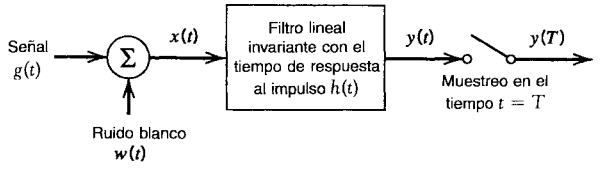


FIGURA 4.1 Receptor lineal.

donde $|g_o(T)|^2$ es la potencia instantánea en la señal de salida, E es el operador de esperanza estadística y $E[n^2(t)]$ es una medida de la potencia promedio del ruido de salida. El requerimiento implica especificar la respuesta al impulso $h(t)$ del filtro, de manera que la relación señal a ruido de salida en la ecuación (4.3) se haga máxima.

Sea $G(f)$ la transformada de Fourier de la señal conocida $g(t)$, y $H(f)$ la respuesta en frecuencia del filtro. En ese caso, la transformada de Fourier de la señal de salida $g_o(t)$ es igual a $H(f)G(f)$ y $g_o(t)$ está determinada por la transformada inversa de Fourier

$$g_o(t) = \int_{-\infty}^{\infty} H(f)G(f)\exp(j2\pi ft)\, df \tag{4.4}$$

Por tanto, cuando la salida del filtro se muestrea en el tiempo $t = T$, tenemos (a falta de ruido del canal)

$$|g_o(T)|^2 = \left|\int_{-\infty}^{\infty} H(f)G(f)\exp(j2\pi fT)\, df\right|^2 \tag{4.5}$$

Considere a continuación el efecto de la salida del filtro por causa del ruido $w(t)$ actuando solo. La densidad espectral de potencia $S_N(f)$ del ruido de salida $n(t)$ es igual a la densidad espectral de potencia del ruido de entrada $w(t)$ multiplicado por la respuesta en magnitud al cuadrado $|H(f)|^2$ (vea la sección 1.7). Puesto que $w(t)$ es blanco con densidad espectral de potencia constante $N_0/2$, se concluye que

$$S_N(f) = \frac{N_0}{2}|H(f)|^2 \tag{4.6}$$

La potencia promedio del ruido de salida $n(t)$ es entonces

$$\begin{aligned} E[n^2(t)] &= \int_{-\infty}^{\infty} S_N(f)\, df \\ &= \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2\, df \end{aligned} \tag{4.7}$$

Al sustituir las ecuaciones (4.5) y (4.7) en la (4.3), es posible reescribir la expresión para la relación señal a ruido del pulso pico como

$$\eta = \frac{\left|\int_{-\infty}^{\infty} H(f)G(f)\exp(j2\pi fT)\, df\right|^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2\, df} \tag{4.8}$$

Nuestro problema es encontrar, para una $G(f)$ determinada, la forma particular de la respuesta en frecuencia $H(f)$ del filtro que hace máxima η. Para encontrar la solución a este problema de optimización, aplicamos un resultado matemático conocido como la desigualdad de Schwarz al numerador de la ecuación (4.8).

La deducción de la *desigualdad de Schwarz* se presenta en el capítulo 5. Por ahora es suficiente afirmar que si tenemos dos funciones complejas $\phi_1(x)$ y $\phi_2(x)$ de variable real x, que satisfagan las condiciones

$$\int_{-\infty}^{\infty} |\phi_1(x)|^2\, dx < \infty$$

y

$$\int_{-\infty}^{\infty} |\phi_2(x)|^2\, dx < \infty$$

entonces es posible escribir

$$\left|\int_{-\infty}^{\infty} \phi_1(x)\phi_2(x)\,dx\right|^2 \le \int_{-\infty}^{\infty} |\phi_1(x)|^2\,dx \int_{-\infty}^{\infty} |\phi_2(x)|^2\,dx \tag{4.9}$$

La igualdad en la ecuación (4.9) se cumple si, y sólo si, tenemos

$$\phi_1(x) = k\phi_2^*(x) \tag{4.10}$$

donde k es una constante arbitraria, y el asterisco indica la conjugación compleja.

Volviendo al problema presente, vemos de inmediato que al apelar a la desigualdad de Schwarz (4.9) y al dejar que $\phi_1(x) = H(f)$ y $\phi_2(x) = G(f)\exp(j\pi fT)$, el numerador en la ecuación (4.8) puede reescribirse como

$$\left|\int_{-\infty}^{\infty} H(f)G(f)\exp(j2\pi fT)\,df\right|^2 \le \int_{-\infty}^{\infty} |H(f)|^2\,df \int_{-\infty}^{\infty} |G(f)|^2\,df \tag{4.11}$$

Utilizando esta relación en la ecuación (4.8), se redefine la relación señal a ruido del pulso pico como

$$\eta \le \frac{2}{N_0}\int_{-\infty}^{\infty} |G(f)|^2\,df \tag{4.12}$$

El segundo miembro de esta relación no depende de la respuesta en frecuencia $H(f)$ del filtro sino sólo de la energía de la señal y de la densidad espectral de potencia del ruido. En consecuencia, la relación señal a ruido del pulso pico η será un máximo cuando $H(f)$ se elija de modo que se cumpla la igualdad; esto es,

$$\eta_{\text{máx}} = \frac{2}{N_0}\int_{-\infty}^{\infty} |G(f)|^2\,df \tag{4.13}$$

En forma correspondiente, $H(f)$ toma su valor óptimo denotado por $H_{\text{opt}}(f)$. Para encontrar este valor óptimo utilizamos la ecuación (4.10), la cual, en la situación presente, produce

$$H_{\text{opt}}(f) = kG^*(f)\exp(-j2\pi fT) \tag{4.14}$$

donde $G^*(f)$ es la compleja conjugada de la transformada de Fourier de la señal de entrada $g(t)$, y k es un factor de escalamiento de dimensiones apropiadas. Esta relación establece que, excepto por el factor $k\exp(-j2\pi fT)$, la respuesta en frecuencia del filtro óptimo es la misma que la de la compleja conjugada de la transformada de Fourier de la señal de entrada.

La ecuación (4.14) especifica el filtro óptimo en el dominio de la frecuencia. Para caracterizarlo en el dominio del tiempo, tomamos la transformada inversa de Fourier de $H_{\text{opt}}(f)$ en la ecuación (4.14) para obtener la respuesta al impulso del filtro óptimo como

$$h_{\text{opt}}(t) = k\int_{-\infty}^{\infty} G^*(f)\exp[-j2\pi f(T-t)]\,df \tag{4.15}$$

Puesto que para una señal real $g(t)$ tenemos $G^*(f) = G(-f)$, es posible reescribir la ecuación (4.15) como

$$\begin{aligned} h_{\text{opt}}(t) &= k\int_{-\infty}^{\infty} G(-f)\exp[-j2\pi f(T-t)]\,df \\ &= k\int_{-\infty}^{\infty} G(f)\exp[j2\pi f(T-t)]\,df \\ &= kg(T-t) \end{aligned} \tag{4.16}$$

La ecuación (4.16) muestra que la respuesta al impulso del filtro óptimo, salvo por el factor de escalamiento k, es una versión invertida y retardada en el tiempo de la señal de entrada $g(t)$; esto es, se "acopla" con la señal de entrada. Un filtro lineal invariante con el tiempo definido de

esta manera recibe el nombre de filtro de acoplamiento. Advierta que al obtener el filtro de acoplamiento, la única suposición que hemos hecho acerca del ruido de entrada $w(t)$ es que éste es estacionario y blanco con media cero y densidad espectral de potencia $N_0/2$. En otras palabras, no se hizo ninguna suposición acerca de la estadística del ruido del canal $w(t)$.

Propiedades de los filtros de acoplamiento

Advertimos que un filtro, el cual se acopla a una señal de pulso $g(t)$ de duración T, se caracteriza por una respuesta al impulso que es una versión invertida y retardada en el tiempo de la entrada $g(t)$, como indica

$$h_{opt}(t) = kg(T - t)$$

En otras palabras, la respuesta al impulso $h_{opt}(t)$ está definida en forma única, salvo por el retardo T y el factor de escalamiento k, por la forma de onda de la señal de pulso $g(t)$ al cual se acopla el filtro. En el dominio de la frecuencia, el filtro de acoplamiento se caracteriza por una respuesta en frecuencia que es, excepto por el factor de retraso, la compleja conjugada de la transformada de Fourier de la entrada $g(t)$, de la forma que indica

$$H_{opt}(f) = kG^*(f)\exp(-j2\pi fT)$$

El resultado más importante en el cálculo del desempeño de sistemas de procesamiento de señales que utilizan filtros de acoplamiento, es quizá el siguiente:

La relación señal a ruido del pulso pico de un filtro de acoplamiento depende sólo de la proporción entre la energía de la señal y la densidad espectral de potencia del ruido blanco a la entrada del filtro.

Para demostrar esta propiedad, considere un filtro acoplado a una señal conocida $g(t)$. La transformada de Fourier de la salida $g_o(t)$ que resulta del filtro de acoplamiento es

$$\begin{aligned} G_o(f) &= H_{opt}(f)G(f) \\ &= kG^*(f)G(f)\exp(-j2\pi fT) \\ &= k|G(f)|^2\exp(-j2\pi fT) \end{aligned} \tag{4.17}$$

Al utilizar la ecuación (4.17) en la fórmula para la transformada inversa de Fourier, encontramos que la salida del filtro de acoplamiento en el tiempo $t = T$ es

$$\begin{aligned} g_o(T) &= \int_{-\infty}^{\infty} G_o(f)\exp(j2\pi fT)\,df \\ &= k\int_{-\infty}^{\infty} |G(f)|^2\,df \end{aligned}$$

De acuerdo con el *teorema de energía de Rayleigh, la integral del espectro de magnitud al cuadrado de una señal de pulso con respecto a la frecuencia es igual a la energía E de la señal:*

$$E = \int_{-\infty}^{\infty} g^2(t)dt = \int_{-\infty}^{\infty} |G(f)|^2\,df$$

Por tanto,

$$g_o(T) = kE \tag{4.18}$$

Al sustituir la ecuación (4.14) en la (4.7), encontramos que la potencia promedio del ruido de salida es

$$\begin{aligned} E[n^2(t)] &= \frac{k^2N_0}{2}\int_{-\infty}^{\infty} |G(f)|^2\,df \\ &= k^2N_0E/2 \end{aligned} \tag{4.19}$$

donde de nuevo hemos utilizado el teorema de energía de Rayleigh. Por tanto, la relación señal a ruido del pulso pico tiene el valor máximo

$$\eta_{\text{máx}} = \frac{(kE)^2}{(k^2 N_0 E/2)} = \frac{2E}{N_0} \tag{4.20}$$

De la ecuación (4.20) vemos que la dependencia de la entrada $g(t)$ con respecto a la forma de onda se ha eliminado por completo por medio del filtro de acoplamiento. En consecuencia, al evaluar la capacidad de un receptor de filtro de acoplamiento para combatir el ruido blanco aditivo, observamos que todas las señales que tienen la misma energía son igualmente efectivas. Advierta que la energía E de la señal está en joules y que la densidad espectral del ruido $N_0/2$ se determina en watts por hertz, de modo que la proporción $2E/N_0$ es adimensional; sin embargo, las dos cantidades tienen diferente significado físico. Nos referimos a E/N_0 como la *relación energía de la señal a la densidad espectral de ruido*.

EJEMPLO 4.1 Filtro de acoplamiento para un pulso rectangular

Considere la señal $g(t)$ en la forma de un pulso rectangular de amplitud A y duración T, como se muestra en la figura 4.2*a*. En este ejemplo, la respuesta al impulso $h(t)$ del filtro de acoplamiento tiene exactamente la misma forma de onda de la propia señal. La señal de salida $g_o(t)$ del filtro de acoplamiento que se produce en respuesta a la señal de entrada $g(t)$ tiene una forma de onda triangular (figura 4.2*b*).

El valor máximo de la señal de salida $g_o(t)$ es igual a kA^2T, que es la energía de la señal de entrada $g(t)$ escalada por el factor k; este valor máximo ocurre en $t = T$, como se indica en la figura 4.2*b*.

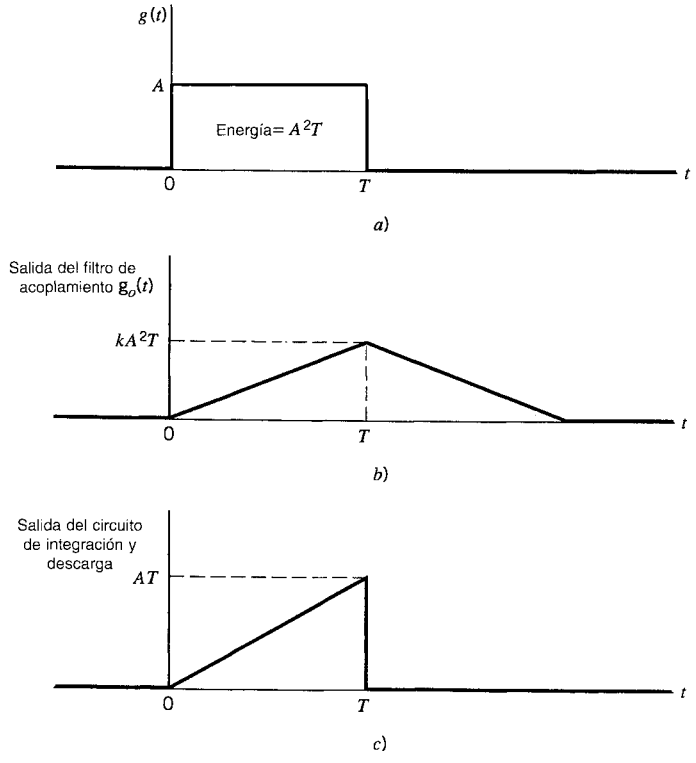


FIGURA 4.2 *a*) Pulso rectangular. *b*) Salida del filtro de acoplamiento. *c*) Salida del integrador.

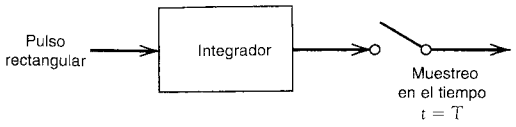


FIGURA 4.3 Circuito de integración y descarga.

Para el caso especial de un pulso rectangular, el filtro de acoplamiento puede ponerse en práctica utilizando el *circuito de integración y descarga*, del cual se presenta un diagrama de bloques en la figura 4.3. El integrador calcula el área bajo el pulso rectangular, y la salida resultante se muestrea después en el tiempo $t = T$, donde T es la duración del pulso. Inmediatamente después de $t = T$, el integrador regresa a su condición inicial; de ahí el nombre del circuito. La figura 4.2c muestra la forma de onda de salida del circuito de integración y descarga para el pulso rectangular de la figura 4.2*a*. Vemos que para $0 \leq t \leq T$, la salida de este circuito tiene la *misma forma de onda* que la que aparece a la salida del filtro de acoplamiento; la diferencia en la notación que se utiliza para describir sus valores pico no tiene importancia práctica. ◀

4.3 *Tasa de error debido al ruido*

En la sección 3.8 presentamos una discusión cualitativa del efecto del ruido de canal en el desempeño de un sistema PCM binario. Ahora que contamos con el filtro de acoplamiento como el detector óptimo de un pulso conocido en ruido blanco aditivo, estamos listos para obtener una fórmula relativa a la tasa de error que origina el ruido en un sistema de este tipo.

Para proceder con el análisis, considere un sistema PCM binario basado en la *transmisión de señales polares sin retorno a cero* (NRZ). En esta forma de transmisión de señales los símbolos 1 y 0 se representan por medio de pulsos rectangulares positivos y negativos de igual amplitud y duración. El ruido del canal se modela como *ruido blanco gaussiano aditivo* $w(t)$ de media cero y densidad espectral de potencia $N_0/2$; la suposición de ruido gaussiano se necesita en los cálculos posteriores. En el intervalo de transmisión $0 \leq t \leq T_b$, la señal recibida se escribe consecuentemente como:

$$x(t) = \begin{cases} +A + w(t), & \text{se envió el símbolo 1} \\ -A + w(t), & \text{se envió el símbolo 0} \end{cases} \tag{4.21}$$

donde T_b es la *duración de bit* y A es la *amplitud de pulso transmitida*. Se supone que el receptor ha adquirido conocimiento de los tiempos de inicio y término de cada pulso transmitido; dicho de otra manera, el receptor cuenta con conocimiento previo de la forma del pulso pero no de su polaridad. Dada la señal de ruido $x(t)$ se requiere que el receptor tome una decisión en cada intervalo de transmisión de señal en cuanto a si el símbolo transmitido es un 1 o un 0.

La estructura del receptor que se utiliza para llevar a cabo este proceso de toma de decisiones se ilustra en la figura 4.4. Consiste en un filtro de acoplamiento seguido por un muestreador

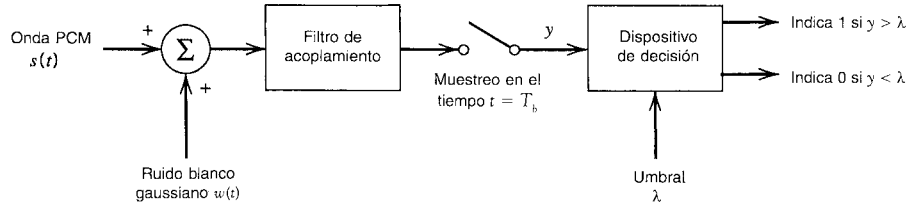


FIGURA 4.4 Receptor para la transmisión de banda base de una onda PCM de codificación binaria utilizando transmisión de señales polares sin retorno a cero.

y finalmente por el dispositivo de decisión. El filtro se acopla a un pulso rectangular de amplitud A y duración T_b, aprovechando la información de temporización de bits disponible para el receptor. La salida del filtro de acoplamiento que se origina se muestrea al final de cada intervalo de transmisión de señales. La presencia de ruido del canal $w(t)$ añade aleatoriedad a la salida del filtro de acoplamiento.

Sea y el valor de muestreo obtenido al final del intervalo de transmisión de la señal. El valor de muestreo y se compara con un *umbral* λ preestablecido en el dispositivo de decisión. Si se supera el umbral, el receptor toma la decisión en favor del símbolo 1; si no, la decisión se hace en favor del símbolo 0. Adoptamos la convención de que cuando el valor de muestreo y es exactamente igual al umbral λ, el receptor sólo supone cuál símbolo se transmitió; una decisión de estas características es la misma que se obtiene al lanzar una moneda sin alterar, el resultado de lo cual no afectará la probabilidad promedio del error.

Existen dos tipos posibles de error que considerar:

1. El símbolo 1 se elige cuando realmente se transmitió un 0; nos referiremos al error anterior como un *error de primera clase*.
2. El símbolo 0 se elige cuando en realidad se transmitió un 1; nos referiremos a este error como un *error de segunda clase*.

Para determinar la probabilidad promedio del error, consideramos estas dos situaciones por separado.

Suponga que se envió el símbolo 0, entonces, de acuerdo con la ecuación (4.21), la señal recibida es

$$x(t) = -A + w(t), \quad 0 \le t \le T_b \tag{4.22}$$

En forma correspondiente, la salida del filtro de acoplamiento, muestreada al tiempo $t = T_b$, está dada por (de acuerdo con el ejemplo 4.1 con kAT_b igualada a la unidad por conveniencia en la presentación)

$$\begin{aligned} y &= \int_0^{T_b} x(t)\,dt \\ &= -A + \frac{1}{T_b}\int_0^{T_b} w(t)\,dt \end{aligned} \tag{4.23}$$

que representa el valor de muestreo de una variable aleatoria Y. Por el hecho de que el ruido $w(t)$ es blanco y gaussiano, es posible caracterizar la variable aleatoria Y del modo siguiente:

- La variable aleatoria Y presenta una distribución gaussiana con media $-A$.
- La varianza de la variable aleatoria Y es

$$\begin{aligned} \sigma_Y^2 &= E\left[(Y + A)^2\right] \\ &= \frac{1}{T_b^2} E\left[\int_0^{T_b}\int_0^{T_b} w(t)w(u)\,dt\,du\right] \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} E\left[w(t)w(u)\right]dt\,du \\ &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b} R_W(t,u)\,dt\,du \end{aligned} \tag{4.24}$$

donde $R_W(t, u)$ es la función de autocorrelación del ruido blanco $w(t)$. Puesto que $w(t)$ es blanco con una densidad espectral de potencia $N_0/2$, tenemos

$$R_W(t,u) = \frac{N_0}{2}\delta(t-u) \tag{4.25}$$

donde $\delta(t-u)$ es una función delta recorrida en el tiempo. De tal modo, la sustitución de la ecuación (4.25) en la (4.24) produce

$$\begin{aligned}\sigma_Y^2 &= \frac{1}{T_b^2}\int_0^{T_b}\int_0^{T_b}\frac{N_0}{2}\delta(t-u)\,dt\,du \\ &= \frac{N_0}{2T_b}\end{aligned} \tag{4.26}$$

donde hemos utilizado la propiedad de filtrado de la función delta y el hecho de que su área es la unidad. La función densidad de probabilidad condicional de la variable aleatoria Y, puesto que se envió el símbolo 0, es entonces

$$f_Y(y\,|\,0) = \frac{1}{\sqrt{\pi N_0/T_b}}\exp\left(-\frac{(y+A)^2}{N_0/T_b}\right) \tag{4.27}$$

Esta función se grafica en la figura 4.5*a*. Considere que p_{10} denota la *probabilidad condicional de error, dado que se envió el símbolo* 0. Esta probabilidad se define por medio del área sombreada bajo la curva de $f_Y(y\,|\,0)$ desde el umbral λ hasta infinito, lo cual corresponde al intervalo de valores supuestos para y relativa a una decisión en favor del símbolo 1. A falta de ruido, la salida del filtro de acoplamiento y muestreada en el tiempo $t = T_b$ es igual a $-A$. Cuando el ruido está presente, y asume ocasionalmente un valor mayor que λ, en cuyo caso se produce un error. La probabilidad de este error, condicional al envío del símbolo 0, está definida por

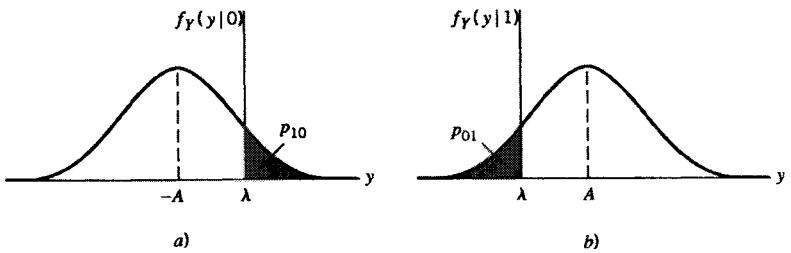
$$\begin{aligned}p_{10} &= P(y > \lambda \mid \text{se envió el símbolo 0}) \\ &= \int_\lambda^\infty f_Y(y|0)\,dy \\ &= \frac{1}{\sqrt{\pi N_0/T_b}}\int_\lambda^\infty \exp\left(-\frac{(y+A)^2}{N_0/T_b}\right)dy\end{aligned} \tag{4.28}$$

En este punto de la exposición nos apartamos un poco e introducimos la definición de la llamada *función de error complementaria*:[3]

$$\operatorname{ferc}(u) = \frac{2}{\sqrt{\pi}}\int_u^\infty \exp(-z^2)\,dz \tag{4.29}$$

que se relaciona estrechamente con la distribución gaussiana. Para valores grandes y positivos de u tenemos la siguiente *cota superior* en la función de error complementaria:

$$\operatorname{ferc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi u}} \tag{4.30}$$



a) *b)*

FIGURA 4.5 Análisis de ruido del sistema PCM. *a)* Función de densidad de probabilidad de la variable aleatoria Y a la salida del filtro de acoplamiento cuando se transmite 0. *b)* Función de densidad de probabilidad de Y cuando se transmite 1.

Para reformular esta probabilidad condicional de error p_{10} en términos de la función de error complementaria, definimos primero una nueva variable

$$z = \frac{y + A}{\sqrt{N_0 / T_b}}$$

En consecuencia, es posible reescribir la ecuación (4.28) en la forma compacta

$$\begin{aligned} p_{10} &= \frac{1}{\sqrt{\pi}} \int_{(A+\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2} \operatorname{ferc}\left(\frac{A + \lambda}{\sqrt{N_0 / T_b}} \right) \end{aligned} \tag{4.31}$$

Se supone a continuación que se transmitió el símbolo 1. Esta vez la variable aleatoria gaussiana Y representada por el valor de muestreo y de la salida del filtro de acoplamiento tiene una media $+A$ y varianza $N_0/2T_b$. Advierta que, en comparación con la situación en la que se envía el símbolo 0, la media de la variable aleatoria Y ha cambiado, pero su varianza es exactamente la misma que antes. La función de densidad de la probabilidad condicional de Y, puesto que se envió el símbolo 1, es entonces

$$f_Y(y|1) = \frac{1}{\sqrt{\pi N_0 / T_b}} \exp\left(-\frac{(y - A)^2}{N_0 / T_b} \right) \tag{4.32}$$

la cual se grafica en la figura 4.5*b*. Sea p_{01} la *probabilidad condicional de error, debido a que se envió el símbolo 1*. Esta probabilidad se define por medio del área sombreada bajo la curva $f_Y(y|1)$ extendiéndose desde $-\infty$ hasta el umbral λ, lo cual corresponde al intervalo de valores asumidos por y para una decisión en favor del símbolo 0. A falta de ruido, la salida y del filtro de acoplamiento muestreada en el tiempo $t = T_b$ es igual a $+A$. Cuando está presente el ruido, y toma ocasionalmente un valor menor que λ, y se realiza un error. La probabilidad de este error, condicional al envío del símbolo 1, está definida por

$$\begin{aligned} p_{01} &= P(y < \lambda \mid \text{se envió el símbolo 1}) \\ &= \int_{-\infty}^{\lambda} f_Y(y|1)\, dy \\ &= \frac{1}{\sqrt{\pi N_0 / T_b}} \int_{-\infty}^{\lambda} \exp\left(-\frac{(y - A)^2}{N_0 / T_b} \right) dy \end{aligned} \tag{4.33}$$

Para expresar p_{01} en términos de la función de error complementaria, definimos esta vez una nueva variable

$$z = \frac{A - y}{\sqrt{N_0 / T_b}}$$

Por tanto, podemos reformular la ecuación (4.33) en la forma compacta

$$\begin{aligned} p_{01} &= \frac{1}{\sqrt{\pi}} \int_{(A-\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz \\ &= \frac{1}{2} \operatorname{ferc}\left(\frac{A - \lambda}{\sqrt{N_0 / T_b}} \right) \end{aligned} \tag{4.34}$$

Habiendo determinado las probabilidades condicionales del error, p_{10} y p_{01}, nuestra siguiente tarea es obtener la fórmula para la *probabilidad promedio del error de símbolo*, denotada por P_e. Aquí advertimos que estos dos tipos posibles de error son eventos mutuamente excluyentes,

de manera que si el receptor, en un instante de muestreo particular, escoge el símbolo 1, entonces se excluye la aparición del símbolo 0, y viceversa. Considere que p_0 y p_1 denotan las probabilidades *a priori* de transmitir los símbolos 0 y 1, respectivamente. Por tanto, la *probabilidad promedio del error de símbolo* P_e en el receptor está dada por

$$\begin{aligned} P_e &= p_0 p_{10} + p_1 p_{01} \\ &= \frac{p_0}{2}\operatorname{ferc}\left(\frac{A+\lambda}{\sqrt{N_0/T_b}}\right) + \frac{p_1}{2}\operatorname{ferc}\left(\frac{A-\lambda}{\sqrt{N_0/T_b}}\right) \end{aligned} \tag{4.35}$$

De la ecuación (4.35) vemos que P_e es en realidad una función del umbral λ, lo cual sugiere de inmediato la necesidad de formular un *umbral óptimo* que minimice P_e. Para dicha optimización utilizamos la *regla de Leibniz*.

Considere la integral

$$\int_{a(u)}^{b(u)} f(z,u)\, dz$$

La regla de Leibniz establece que la derivada de esta integral con respecto a u es

$$\frac{d}{du}\int_{a(u)}^{b(u)} f(z,u)dz = f(b(u),u)\frac{db(u)}{du} - f(a(u),u)\frac{da(u)}{du} + \int_{a(u)}^{b(u)} \frac{\partial f(z,u)}{\partial u}\, dz$$

En el problema presente, observamos de la definición de la función de error complementaria en la ecuación (4.29) que

$$\begin{aligned} f(z,u) &= \frac{2}{\sqrt{\pi}}\exp(-z^2) \\ a(u) &= u \\ b(u) &= \infty \end{aligned}$$

La aplicación de la regla de Leibniz a la función de error complementaria produce entonces

$$\frac{d}{du}\operatorname{ferc}(u) = -\frac{1}{\sqrt{\pi}}\exp(-u^2) \tag{4.36}$$

De este modo, al diferenciar la ecuación (4.35) con respecto a λ haciendo uso de la fórmula de la ecuación (4.36), e igualar a cero y simplificar términos, obtenemos el umbral óptimo en la forma

$$\lambda_{opt} = \frac{N_0}{4AT_b}\log\left(\frac{p_0}{p_1}\right) \tag{4.37}$$

En el caso especial cuando los símbolos 1 y 0 son igualmente probables, tenemos

$$p_1 = p_0 = \frac{1}{2}$$

en cuyo caso la ecuación (4.37) se reduce a

$$\lambda_{opt} = 0$$

Este resultado se satisface intuitivamente, puesto que establece que en la transmisión de símbolos binarios igualmente probables, debemos elegir el umbral en el punto medio entre las alturas de pulso $-A$ y $+A$ que representan a los dos símbolos 0 y 1. Note que en este caso especial tenemos también

$$p_{01} = p_{10}$$

Un canal para el que las probabilidades condicionales de error p_{01} y p_{10} son iguales se dice que será *simétrico binario*. En forma correspondiente, la probabilidad promedio del error de símbolo en la ecuación (4.35) se reduce a

$$P_e = \frac{1}{2}\text{ferc}\left(\frac{A}{\sqrt{N_0/T_b}}\right) \tag{4.38}$$

Ahora bien, la *energía de la señal transmitida* por bit está definida por

$$E_b = A^2 T_b \tag{4.39}$$

En consecuencia, es posible formular finalmente la probabilidad promedio del error de símbolo para el receptor de la figura 4.4 como

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{4.40}$$

la cual muestra que la *probabilidad promedio del error de símbolo en un canal simétrico binario depende exclusivamente de* E_b/N_0, *la proporción entre la energía de la señal transmitida por bit y la densidad espectral del ruido.*

Utilizando la cota superior de la ecuación (4.30) en la función de error complementaria, es posible acotar de manera correspondiente la probabilidad promedio del error de símbolo para el receptor PCM en la forma

$$P_e < \frac{\exp(-E_b/N_0)}{2\sqrt{\pi E_b/N_0}} \tag{4.41}$$

El receptor PCM de la figura 4.4 exhibe entonces una mejora *exponencial* en la probabilidad promedio del error de símbolo con el aumento en E_b/N_0.

Este importante resultado se ilustra de modo adicional en la figura 4.6, donde la probabilidad promedio del error de símbolo P_e se grafica en función del cociente adimensional E_b/N_0. En particular, vemos que P_e disminuye muy rápido cuando el cociente E_b/N_0 está aumentando, por lo que a la larga un "incremento muy pequeño" en la energía de la señal transmitida hará que la recepción de los pulsos binarios casi no tenga errores, como se expuso en la sección 3.8. Sin embargo, advierta que en términos prácticos el incremento en la energía de la señal tiene que considerarse en el contexto de la polarización; por ejemplo, un aumento de 3-dB en E_b/N_0 es

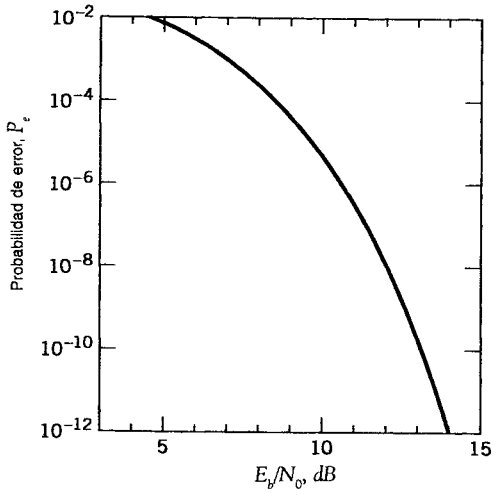


FIGURA 4.6 Probabilidad de error en un receptor PCM.

mucho más fácil de poner en práctica cuando E_b tiene un valor pequeño que cuando su valor es varios órdenes de magnitud mayor.

4.4 *Interferencia de intersímbolo*

La siguiente fuente de errores de bit en un sistema de transmisión de pulsos por la banda base que deseamos estudiar es la interferencia de intersímbolo (ISI), la cual se presenta cuando el canal de comunicación es *dispero*. Sin embargo, antes que nada necesitamos plantear una pregunta fundamental: dada una forma de pulso de interés, ¿cómo utilizarla para transmitir datos en forma M-aria? La respuesta se encuentra en el uso de la *modulación por pulsos discreta*, en la cual la amplitud, duración o posición de los pulsos transmitidos varía de manera discreta de acuerdo con la secuencia de datos dada. Sin embargo, en la transmisión de banda base de datos digitales, el uso de la *modulación por amplitud de pulsos* (PAM) *discreta* constituye uno de los esquemas más eficientes en términos de la utilización de potencia y ancho de banda. En consecuencia, centramos nuestra atención en los sistemas PAM discretos. Empezamos el estudio considerando primero el caso de datos binarios; más adelante en el capítulo abordamos el caso más general de datos M-arios.

Considere entonces un *sistema PAM binario de banda base*, una forma genérica del cual se muestra en la figura 4.7. La secuencia binaria entrante $\{b_k\}$ consiste en los símbolos 1 y 0, cada uno de duración T_b. El *modulador de amplitud por pulsos* modifica esta secuencia binaria en una nueva secuencia de pulsos cortos (aproximándose al impulso unitario), cuya amplitud a_k se representa en la forma polar

$$a_k = \begin{cases} +1 & \text{si el símbolo } b_k \text{ es 1} \\ -1 & \text{si el símbolo } b_k \text{ es 0} \end{cases} \tag{4.42}$$

La secuencia de pulsos cortos producida de este modo se aplica a un *filtro de transmisión* de respuesta al impulso $g(t)$, produciéndose la señal transmitida

$$s(t) = \sum_k a_k g(t - kT_b) \tag{4.43}$$

La señal $s(t)$ se modifica como resultado de la transmisión por el *canal* de respuesta al impulso $h(t)$. Además, el canal agrega ruido aleatorio a la señal a la entrada del receptor. La señal ruidosa $x(t)$ se hace pasar luego por un *filtro de recepción* de respuesta al impulso $c(t)$. La salida $y(t)$ que produce el filtro se muestrea *sincronizadamente* con el transmisor, determinándose los instantes de muestreo por medio de un *reloj* o una señal de *temporización* que suele extraerse de la salida del filtro de recepción. Por último, la secuencia de muestras obtenidas de esa manera se utiliza para reconstruir la secuencia de datos original por medio de un *dispositivo de decisión*. Específicamente, la amplitud de cada muestra se compara con un *umbral* λ. Si se supera el umbral λ, se toma una decisión en favor del símbolo 1. Si ocurre lo contrario, la decisión se toma en favor del símbolo 0.

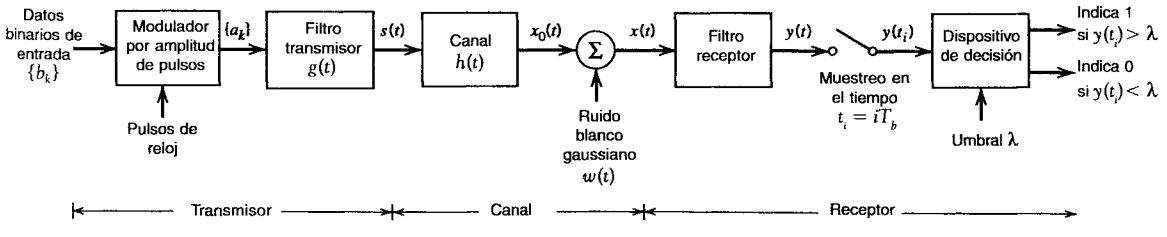


FIGURA 4.7 Sistema de transmisión de datos binarios de la banda base.

Si la amplitud de la muestra es exactamente igual al umbral, el lanzamiento de una moneda determinará cuál símbolo se transmitió (es decir, el receptor simplemente supone).

La salida del filtro de recepción se escribe como

$$y(t) = \mu \sum_k a_k p(t - kT_b) + n(t) \tag{4.44}$$

donde μ es un factor de escalamiento y el pulso $p(t)$ se va a definir. Para ser precisos, es necesario incluir un tiempo de retardo t_0 en el argumento del pulso $p(t - kT_b)$ en la ecuación (4.44) para representar el efecto del retardo de la transmisión a través del sistema. A fin de simplificar la exposición, hemos hecho cero este retardo en la ecuación (4.44) sin perder generalidad.

El pulso escalado $\mu p(t)$ se obtiene mediante una convolución doble que incluye la respuesta al impulso $g(t)$ del filtro de transmisión, la respuesta al impulso $h(t)$ del canal y la respuesta al impulso $c(t)$ del filtro de recepción, como indica

$$\mu p(t) = g(t) \bigstar h(t) \bigstar c(t) \tag{4.45}$$

donde las estrellas indican convolución. Suponemos que el pulso $p(t)$ está *normalizado* fijando

$$p(0) = 1 \tag{4.46}$$

lo cual justifica el empleo de μ como un factor de escalamiento para tener en cuenta los cambios de amplitud en los que se incurre en el curso de la transmisión de la señal a través del sistema.

Puesto que la convolución en el dominio del tiempo se transforma en multiplicación en el dominio de la frecuencia, es posible utilizar la transformada de Fourier para cambiar la ecuación (4.45), en la forma equivalente

$$\mu P(f) = G(f)H(f)C(f) \tag{4.47}$$

donde $P(f)$, $G(f)$, $H(f)$ y $C(f)$ son las transformadas de Fourier de $p(t)$, $g(t)$, $h(t)$ y $c(t)$, respectivamente.

Por último, el término $n(t)$ en la ecuación (4.44) es el ruido que se produce a la salida del filtro de recepción por causa del ruido del canal $w(t)$. Es usual modelar $w(t)$ como un ruido blanco gaussiano de media cero.

La salida del filtro de recepción $y(t)$ se muestrea en el tiempo $t_i = iT_b$ (con i tomando valores enteros), lo que produce [de acuerdo con la ecuación (4.46)]

$$\begin{aligned} y(t_i) &= \mu \sum_{k=-\infty}^{\infty} a_k p[(i-k)T_b] + n(t_i) \\ &= \mu a_i + \mu \sum_{\substack{k=-\infty \\ k \neq i}}^{\infty} a_k p[(i-k)T_b] + n(t_i) \end{aligned} \tag{4.48}$$

En la ecuación (4.48) el primer término μa_i representa la contribución del i-ésimo bit transmitido. El segundo término representa el efecto residual de todos los demás bits transmitidos sobre la decodificación del i-ésimo bit; este efecto residual debido a la ocurrencia de pulsos antes y después del instante de muestreo t_i recibe el nombre de interferencia de intersímbolo (ISI). El último término $n(t_i)$ representa la muestra del ruido en el tiempo t_i.

A falta tanto de interferencia de intersímbolo como de ruido, observamos de la ecuación (4.48) que

$$y(t_i) = \mu a_i$$

lo cual muestra, en estas condiciones ideales, que el i-ésimo bit transmitido se decodifica en forma correcta. La inevitable presencia de la interferencia de intersímbolo y del ruido en el sistema introduce, sin embargo, errores en el dispositivo de recepción a la salida del receptor. Por tanto, en el diseño de los filtros de transmisión y recepción, el objetivo es minimizar los efectos del ruido y de la interferencia de intersímbolo y, por ello, entregar los datos digitales al destinatario con la tasa de error más pequeña posible.

Cuando la relación señal a ruido es alta, como en el caso de un sistema telefónico, por ejemplo, la operación del sistema está bastante limitada por la interferencia de intersímbolo en vez del ruido; en otras palabras, es posible ignorar $n(t_i)$. En el siguiente par de secciones, suponemos que se cumple esta condición, de modo que es posible centrar nuestra atención en la interferencia de intersímbolo y en las técnicas para su control. En particular, el asunto que nos interesa considerar es la determinación de la forma de onda del pulso $p(t)$ para la cual la ISI se elimina completamente.

4.5 *Criterio de Nyquist para la transmisión binaria sin distorsión en banda base*

Por lo común se especifica la respuesta en frecuencia del canal y la forma del pulso transmitido, y el problema radica en determinar las respuestas en frecuencia de los filtros de transmisión y recepción con el fin de reconstruir la secuencia de datos binarios original $\{b_k\}$. El receptor realiza lo anterior *extrayendo* y luego *decodificando* la secuencia correspondiente de coeficientes, $\{a_k\}$, de la salida $y(t)$. La *extracción* implica muestrear la salida $y(t)$ en el tiempo $t = iT_b$. La *decodificación* requiere que la contribución ponderada del pulso $a_k p(iT_b - kT_b)$ para $k = i$ no sufra de ISI por causa de las colas traslapadas de todas las demás contribuciones ponderadas de pulso representadas por $k \neq i$. Esto, a su vez, requiere que *controlemos* el pulso completo $p(t)$, como indica

$$p(iT_b - kT_b) = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \tag{4.49}$$

donde $p(0) = 1$, por normalización. Si $p(t)$ satisface las condiciones de la ecuación (4.49), la salida del receptor $y(t_i)$ dada en la ecuación (4.48) se simplifica en (ignorando el término del ruido)

$$y(t_i) = \mu a_i \quad \text{para toda } i$$

lo cual implica interferencia de intersímbolo igual a cero. Por tanto, las dos condiciones de la ecuación (4.49) aseguran la *recepción perfecta a falta de ruido*.

Desde el punto de vista del diseño, resulta informativo transformar las condiciones de la ecuación (4.49) al dominio de la frecuencia. Considere entonces la secuencia de muestras $\{p(nT_b)\}$, donde $n = 0, \pm 1, \pm 2, \cdots$. Con base en la discusión que se presentó en el capítulo 3 acerca del proceso de muestreo, recordamos que el muestreo en el dominio del tiempo produce periodicidad en el dominio de la frecuencia. En particular, podemos escribir

$$P_\delta(f) = R_b \sum_{n=-\infty}^{\infty} P(f - nR_b) \tag{4.50}$$

donde $R_b = 1/T_b$ es la *tasa de bits* en bits por segundo (b/s); $P_\delta(f)$ es la transformada de Fourier de una secuencia periódica infinita de funciones delta de periodo T_b, cuyas áreas individuales se ponderan por medio de valores de muestreo respectivos de $p(t)$. Esto es, $P_\delta(f)$ está dada por

$$P_\delta(f) = \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} [p(mT_b)\, \delta(t - mT_b)] \exp(-j2\pi ft)\, dt \tag{4.51}$$

Sea el entero $m = i - k$, entonces $i = k$ corresponde a $m = 0$, y de igual modo $i \neq k$ corresponde a $m \neq 0$. Por tanto, al imponer las condiciones de la ecuación (4.49) en los valores de las muestras de $p(t)$ en la integral de la ecuación (4.51), obtenemos

$$\begin{aligned} P_\delta(f) &= \int_{-\infty}^{\infty} p(0)\, \delta(t) \exp(-j2\pi ft)\, dt \\ &= p(0) \end{aligned} \tag{4.52}$$

donde hemos recurrido a la propiedad de filtrado de la función delta. Puesto que según la ecuación (4.46) tenemos $p(0) = 1$, se desprende de las ecuaciones (4.50) y (4.52) que la condición para la interferencia de intersímbolo se satisface si

$$\sum_{n=-\infty}^{\infty} P(f - nR_b) = T_b \tag{4.53}$$

Ahora podemos establecer el *criterio de Nyquist*[4] *para la transmisión de banda base sin distorsión* en la ausencia de ruido: *la función de frecuencia $P(f)$ elimina la interferencia de intersímbolo en muestras que se toman a intervalos T_b siempre que se satisfaga la ecuación (4.53)*. Advierta que $P(f)$ se refiere al sistema completo, incorporando el filtro de transmisión, el canal y el filtro de recepción de acuerdo con la ecuación (4.47).

■ CANAL DE NYQUIST IDEAL

La manera simple de satisfacer la ecuación (4.53) estriba en especificar la función de frecuencia $P(f)$ en la forma de una *función rectangular*, de acuerdo con

$$\begin{aligned} P(f) &= \begin{cases} \dfrac{1}{2W}, & -W < f < W \\ 0, & |f| > W \end{cases} \\ &= \frac{1}{2W}\,\mathrm{rect}\left(\frac{f}{2W}\right) \end{aligned} \tag{4.54}$$

donde $\mathrm{rect}(f)$ representa la *función rectangular* de amplitud unitaria de soporte unitario centrado en $f = 0$, y el ancho de banda completo del sistema W se define mediante

$$W = \frac{R_b}{2} = \frac{1}{2T_b} \tag{4.55}$$

De acuerdo con la solución que describen las ecuaciones (4.54) y (4.55), no se necesita ninguna frecuencia de valor absoluto que exceda la mitad de la tasa de bits. De tal modo, a partir del par 2 de transformadas de Fourier de la tabla A6.3 encontramos que la forma de onda de la señal que produce interferencia de intersímbolo igual a cero está definida por la *función senc*:

$$\begin{aligned} p(t) &= \frac{\mathrm{sen}(2\pi Wt)}{2\pi Wt} \\ &= \mathrm{senc}(2Wt) \end{aligned} \tag{4.56}$$

El valor especial de la tasa de bits $R_b = 2W$ recibe el nombre de *tasa de Nyquist*, y W se denomina el *ancho de banda de Nyquist*. De modo correspondiente, el sistema de transmisión de pulsos por banda base ideal descrito por la ecuación (4.54) en el dominio de la frecuencia o, de manera equivalente, la ecuación (4.56) en el dominio del tiempo, reciben el nombre de *canal de Nyquist ideal*.

Las figuras 4.8*a* y 4.8*b* muestran gráficas de $P(f)$ y $p(t)$, respectivamente. En la figura 4.8*a*, la forma normalizada de la función de frecuencia $P(f)$ se grafica para frecuencias positivas y negativas. En la figura 4.8*b* hemos incluido también los intervalos de transmisión de señales y los correspondientes instantes de muestreo centrados. La función $p(t)$ puede considerarse como la respuesta al impulso de un filtro pasobajas ideal con respuesta en magnitud pasobanda $1/2W$ y ancho de banda W. La función $p(t)$ tiene su valor pico en el origen y pasa por cero en múltiplos enteros de la duración de bit T_b. Es claro que si la forma de onda recibida $y(t)$ se

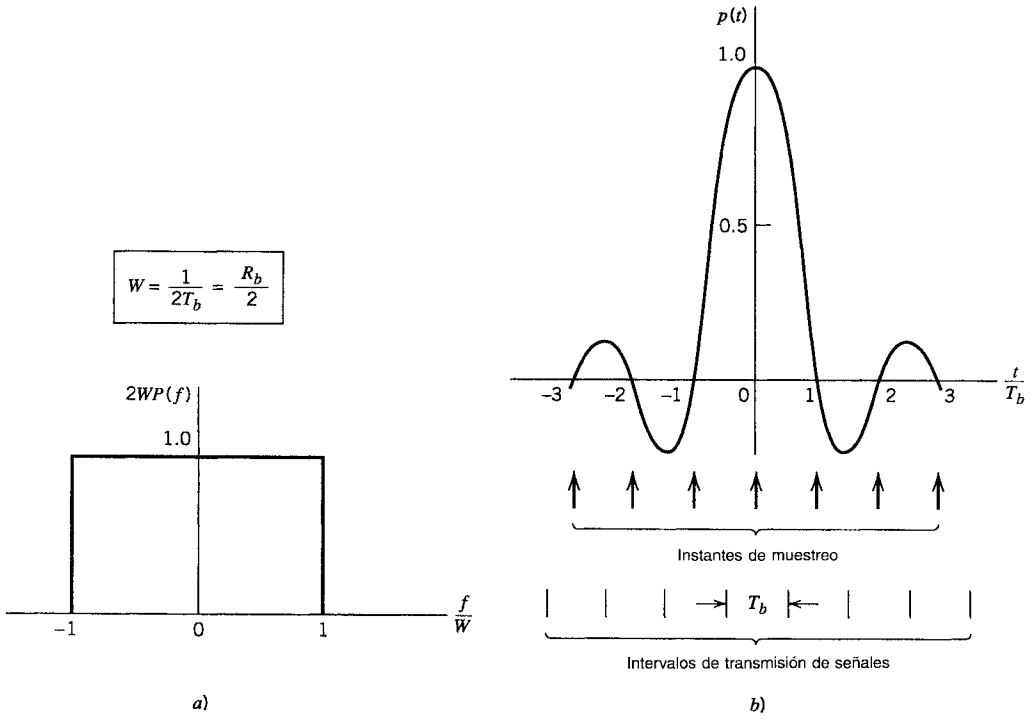


FIGURA 4.8 *a*) Respuesta en magnitud ideal. *b*) Forma del curso básico ideal.

muestrea en los instantes $t = 0, \pm T_b, \pm 2T_b, \cdots$, entonces los pulsos definidos por $\mu p(t - iT_b)$ con amplitud arbitraria μ e índice $i = 0, \pm 1, \pm 2, \cdots$, no interferirán entre sí. Esta condición se ilustra en la figura 4.9 para la secuencia binaria 1011010.

Aunque el uso del canal de Nyquist ideal no logra realmente economía en el ancho de banda de manera que resuelva el problema de interferencia de intersímbolo cero con el mínimo

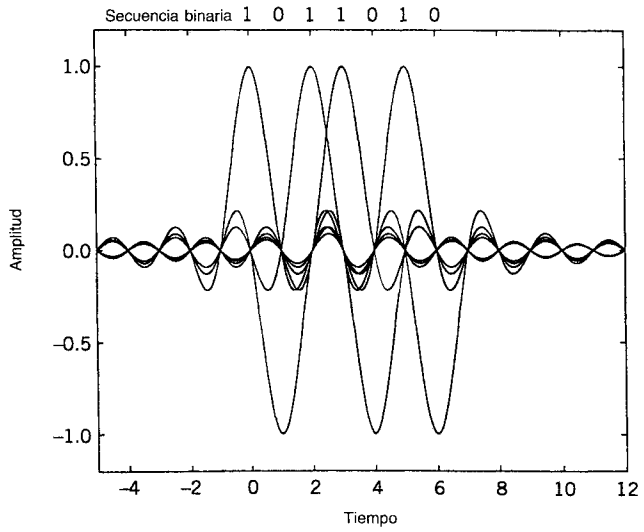


FIGURA 4.9 Una serie de pulsos senc correspondientes a la secuencia 1011010.

de ancho de banda posible, existen dos dificultades prácticas que lo marcan como un objetivo indeseable en el diseño del sistema:

1. Requiere que la característica de la magnitud de $P(f)$ sea plana de $-W$ a W, y cero en cualquier otro lado. Esto es irrealizable físicamente en virtud de las transiciones abruptas en los bordes de la banda $\pm W$.
2. La función $p(t)$ disminuye con $1/|t|$ para $|t|$ grande, originando una baja razón de decaimiento. Lo anterior también es producto de la discontinuidad de $P(f)$ en $\pm W$. En consecuencia, prácticamente no hay margen de error en los tiempos de muestreo en el receptor.

Para evaluar el efecto de este *error de temporización*, considere la muestra de $y(t)$ en $t = \Delta t$, donde Δt es el error de temporización. Para simplificar la exposición podemos igualar a cero el tiempo de muestreo correcto t_i. A falta de ruido tenemos, en consecuencia [de acuerdo con la ecuación (4.48)],

$$\begin{aligned} y(\Delta t) &= \mu \sum_k a_k p(\Delta t - kT_b) \\ &= \mu \sum_k a_k \frac{\operatorname{sen}[2\pi W(\Delta t - kT_b)]}{2\pi W(\Delta t - kT_b)} \end{aligned} \tag{4.57}$$

Puesto que $2WT_b = 1$, por definición, es posible reescribir la ecuación (4.57) en la forma

$$y(\Delta t) = \mu a_0 \operatorname{senc}(2W\,\Delta t) + \frac{\mu \operatorname{sen}(2\pi W\,\Delta t)}{\pi} \sum_{\substack{k \\ k \neq 0}} \frac{(-1)^k a_k}{(2W\,\Delta t - k)} \tag{4.58}$$

El primer término en el segundo miembro de la ecuación (4.58) define el símbolo deseado, en tanto que la serie restante representa la interferencia de intersímbolo provocada por el error de temporización Δt al muestrear la salida $y(t)$. Desafortunadamente, es posible que esta serie diverja, causando por ello decisiones erróneas en el receptor.

ESPECTRO DEL COSENO ELEVADO

Es posible superar las dificultades prácticas que se encontraron con el canal de Nyquist ideal extendiendo el ancho de banda desde el valor mínimo $W = R_b/2$ hasta un valor ajustable entre W y $2W$. Especificamos ahora la respuesta en frecuencia completa $P(f)$ para satisfacer una condición más elaborada que la del canal de Nyquist ideal; específicamente, retenemos tres términos de la ecuación (4.53) y restringimos la banda de frecuencias de interés a $[-W, W]$, como se indica mediante

$$P(f) + P(f - 2W) + P(f + 2W) = \frac{1}{2W}, \quad -W \leq f \leq W \tag{4.59}$$

Podemos idear varias funciones limitadas en banda que satisfagan la ecuación (4.59). Una forma particular de $P(f)$ que incluye muchas características deseables la proporciona el *espectro de coseno elevado*. Esta respuesta en frecuencia consiste en una porción *plana* y una porción *de caída pronunciada* que tiene una forma senoidal como la siguiente:

$$P(f) = \begin{cases} \dfrac{1}{2W}, & 0 \leq |f| < f_1 \\ \dfrac{1}{4W}\left\{1 - \operatorname{sen}\left[\dfrac{\pi(|f| - W)}{2W - 2f_1}\right]\right\}, & f_1 \leq |f| < 2W - f_1 \\ 0, & |f| \geq 2W - f_1 \end{cases} \tag{4.60}$$

El parámetro de frecuencia f_1 y el ancho de banda W se relacionan por medio de

$$\alpha = 1 - \frac{f_1}{W} \tag{4.61}$$

El parámetro α recibe el nombre de *factor de caída*; éste indica el *ancho de banda de exceso* con respecto a la solución ideal, W. De modo específico, el ancho de banda de transmisión B_T se define por medio de

$$\begin{aligned} B_T &= 2W - f_1 \\ &= W(1+\alpha) \end{aligned}$$

La respuesta en frecuencia $P(f)$, normalizada al multiplicarla por $2W$, se grafica en la figura 4.10*a* para tres valores de α, a saber, 0, 0.5 y 1. Vemos que para $\alpha = 0.5$ o 1, la

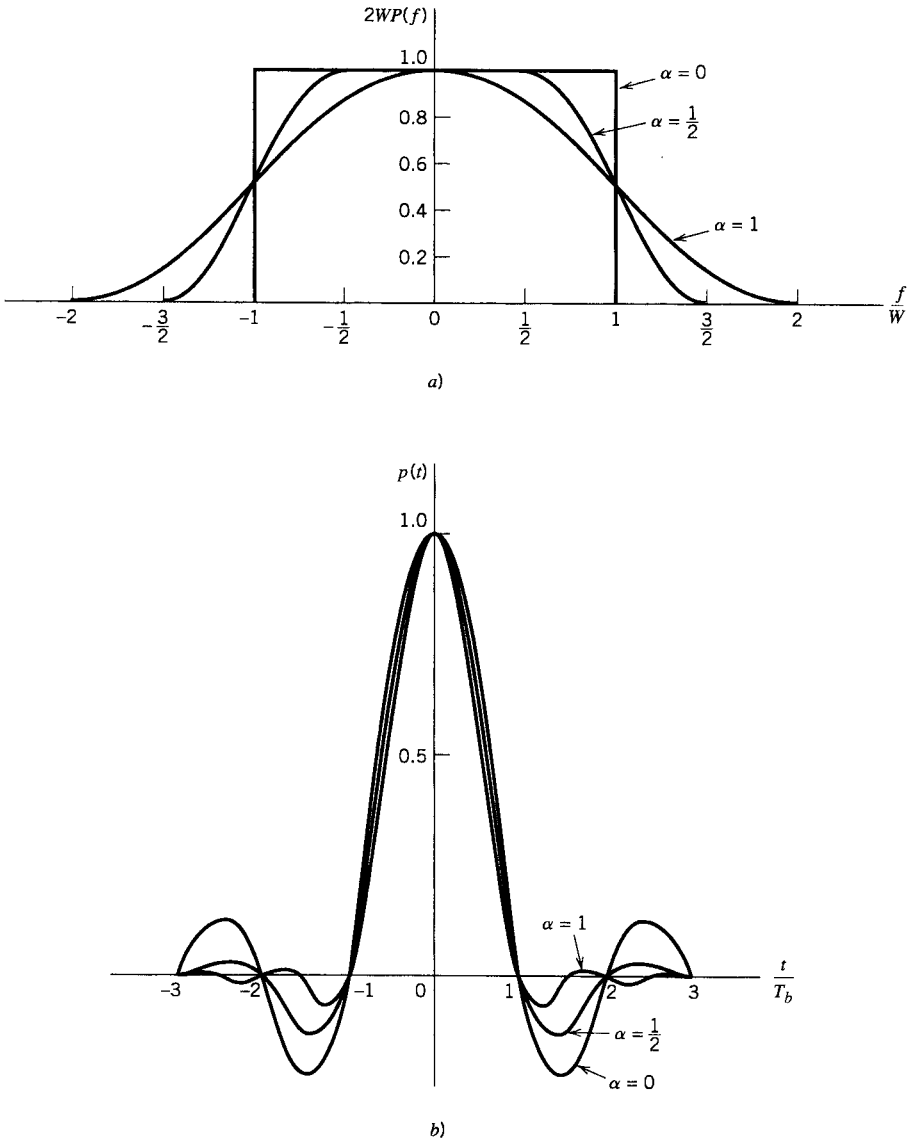


FIGURA 4.10 Respuestas para diferentes factores de caída. *a*) Respuesta en frecuencia. *b*) Respuesta en tiempo.

función $P(f)$ se corta en forma gradual cuando se compara con el canal de Nyquist ideal (es decir, $\alpha = 0$) y es, por tanto, más fácil de ejecutar en la práctica. La función $P(f)$ exhibe también simetría impar con respecto al ancho de banda de Nyquist W, posibilitando la satisfacción de la ecuación (4.59).

La respuesta en tiempo $p(t)$ es la transformada inversa de Fourier de la respuesta en frecuencia $P(f)$. Por tanto, al utilizar la $P(f)$ definida en la ecuación (4.60), obtenemos el resultado (vea el problema 4.13)

$$p(t) = (\text{senc}(2Wt))\left(\frac{\cos(2\pi\alpha Wt)}{1-16\alpha^2W^2t^2}\right) \tag{4.62}$$

el cual se grafica en la figura 4.10*b* para $\alpha = 0$, 0.5 y 1.

La respuesta en tiempo $p(t)$ consiste en el producto de dos factores: el factor $\text{senc}(2Wt)$ que caracteriza al canal de Nyquist ideal y un segundo factor que disminuye como $1/|t|^2$ para $|t|$ grande. El primer factor asegura cruces por cero de $p(t)$ en los instantes de muestreo deseados $t = iT$, con i un entero (positivo y negativo). El segundo factor reduce las colas del pulso bastante abajo de lo que se obtenía a partir del canal de Nyquist ideal, por lo que la transmisión de ondas binarias utilizando pulsos de estas características resulta relativamente insensible a los errores del tiempo de muestreo. De hecho, para $\alpha = 1$ tenemos la caída más gradual en el sentido de que las amplitudes de las colas oscilatorias de $p(t)$ son más pequeñas. En consecuencia, la cantidad de interferencia de intersímbolo que se produce a partir del error de temporización disminuye cuando el factor de caída α se incrementa de cero a la unidad.

El caso especial con $\alpha = 1$ (es decir, $f_1 = 0$) se conoce como la característica de *caída del coseno completo*, para la cual la respuesta en frecuencia de la ecuación (4.60) se simplifica en

$$P(f) = \begin{cases} \dfrac{1}{4W}\left[1+\cos\left(\dfrac{\pi f}{2W}\right)\right], & 0 < |f| < 2W \\ 0, & |f| \geq 2W \end{cases} \tag{4.63}$$

De modo correspondiente, la respuesta en tiempo $p(t)$ se simplifica en

$$p(t) = \frac{\text{senc}(4Wt)}{1-16W^2t^2} \tag{4.64}$$

Esta respuesta en tiempo exhibe dos propiedades interesantes:

1. En $t = \pm T_b/2 = \pm 1/4W$, tenemos $p(t) = 0.5$; esto es, el ancho de pulso medido a la mitad de la amplitud es exactamente igual a la duración de bit T_b.
2. Hay cruces por cero en $t = \pm 3T_b/2, \pm 5T_b/2, \cdots$, además de los usuales cruces por cero en los instantes de muestreo $t = \pm T_b, \pm 2T_b, \cdots$.

Estas dos propiedades son sumamente útiles al extraer una señal de temporización de una señal recibida con fines de sincronización. Sin embargo, el precio que se paga por esta deseable propiedad es el uso de un ancho de banda de canal del doble del que requiere el canal de Nyquist ideal correspondiente a $\alpha = 0$.

▶ EJEMPLO 4.2 Requerimiento de ancho de banda del sistema T1

En el ejemplo 3.2 del capítulo 3 describimos el formato de señal para el sistema de portadora T1 que se usa para multiplexar 24 entradas de voz independientes, con base en una palabra de PCM de 8 bits. Se

demostró que la duración de la señal multiplexada por división de tiempo resultante (incluyendo un bit de trama) es

$$T_b = 0.647 \ \mu s$$

Suponiendo el empleo de un canal de Nyquist ideal, se concluye que el ancho de banda de transmisión mínimo B_T del sistema T1 es (para $\alpha = 0$)

$$B_T = W = \frac{1}{2T_b} = 772 \text{ kHz}$$

Sin embargo, un valor más realista para el ancho de banda de transmisión necesario se obtiene al utilizar la característica de caída de coseno completo con $\alpha = 1$. En este caso, encontramos que

$$B_T = W(1+\alpha) = 2W = \frac{1}{T_b} = 1.544 \text{ MHz}$$

4.6 *Codificación de nivel correlativa*

Hasta ahora hemos considerado la interferencia de intersímbolo como un fenómeno deseable que produce una degradación en el desempeño del sistema. En realidad, su solo nombre connota un efecto perjudicial. A pesar de eso, al agregar de manera controlada interferencia de intersímbolo a la señal transmitida, es posible conseguir una velocidad de transmisión de señales igual a la velocidad de Nyquist de $2W$ símbolos por segundo en un canal de ancho de banda de W hertz. Los esquemas de este tipo reciben el nombre de *codificación de nivel correlativo* o *transmisión de señales de respuesta parcial.*[5] El diseño de estos esquemas se basa en la siguiente premisa: puesto que se conoce la interferencia de intersímbolo que se introduce en la señal transmitida, su efecto puede interpretarse en el receptor de manera determinística. De ese modo, es posible considerar la codificación de nivel correlativo como un método práctico para alcanzar la velocidad de transmisión de señales teórica máxima de $2W$ símbolos por segundo en un ancho de banda de W hertz, según el postulado de Nyquist, utilizando filtros realizables y tolerantes a las perturbaciones.

Transmisión de señales duobinarias

La idea básica de la codificación de nivel correlativo se ilustrará ahora considerando el ejemplo específico de la *transmisión de señales duobinarias*, donde "duo" implica duplicar la capacidad de transmisión de un sistema binario puro. Esta forma particular de codificación de nivel correlativo también se denomina *respuesta parcial de clase I.*

Considere una secuencia de entrada binaria $\{b_k\}$ consistente en símbolos binarios 1 y 0 no correlacionados, cada uno con duración T_b. Como antes, esta secuencia se aplica a un modulador por amplitud de pulsos produciéndose una secuencia de dos niveles de pulsos cortos (aproximándose a un impulso unitario), cuya amplitud a_k está definida por

$$a_k = \begin{cases} +1 & \text{si el símbolo } b_k \text{ es } 1 \\ -1 & \text{si el símbolo } b_k \text{ es } 0 \end{cases} \tag{4.65}$$

Cuando se aplica esta secuencia a un *codificador duobinario*, éste se convierte en una *salida de tres niveles*, a saber, -2, 0 y $+2$. Para producir esta transformación, es posible utilizar el esquema que se muestra en la figura 4.11. La secuencia de dos niveles $\{a_k\}$ se hace pasar primero por un filtro simple que implica un solo elemento de retardo y un sumador. Para cada impulso

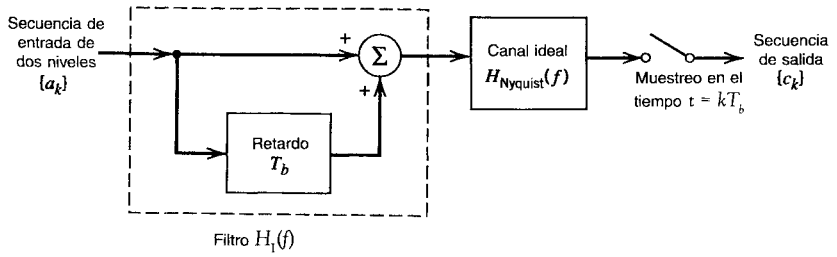


FIGURA 4.11 Esquema de señalización duobinaria.

unitario aplicado a la entrada de este filtro, obtenemos dos impulsos unitarios separados T_b segundos a la salida del filtro. Es posible, por tanto, expresar la salida del codificador duobinario c_k como la suma del pulso de entrada presente a_k y su valor previo a_{k-1}, como indica

$$c_k = a_k + a_{k-1} \tag{4.66}$$

Uno de los efectos de la transformación que describe la ecuación (4.66) es cambiar la secuencia de entrada $\{a_k\}$ de pulsos de dos niveles no correlacionados en una secuencia $\{c_k\}$ de pulsos de tres niveles correlacionados. Esta correlación entre los pulsos adyacentes puede considerarse como la introducción de modo artificial de la interferencia de intersímbolo en la señal transmitida. Sin embargo, la interferencia de intersímbolo que se introduce de esta manera se encuentra bajo el control del diseñador, lo cual constituye la base de la codificación correlativa.

Un elemento de retardo ideal, que produce un retardo T_b, tiene la respuesta en frecuencia $\exp(-j2\pi fT_b)$, por lo que la respuesta en frecuencia del filtro simple de retardo de línea de la figura 4.11 es $1 + \exp(-j2\pi fT_b)$. Por consiguiente, la respuesta en frecuencia completa de este filtro conectado en cascada con un canal de Nyquist ideal es

$$\begin{aligned} H_1(f) &= H_{\text{Nyquist}}(f)[1 + \exp(-j2\pi fT_b)] \\ &= H_{\text{Nyquist}}(f)[\exp(j\pi fT_b) + \exp(-j\pi fT_b)]\exp(-j\pi fT_b) \\ &= 2H_{\text{Nyquist}}(f)\cos(\pi fT_b)\exp(-j\pi fT_b) \end{aligned} \tag{4.67}$$

donde el subíndice I en $H_1(f)$ indica la clase pertinente de respuesta parcial. Para un canal de Nyquist ideal de ancho de banda $W = 1/2T_b$, tenemos (ignorando el factor de escalamiento T_b)

$$H_{\text{Nyquist}}(f) = \begin{cases} 1, & |f| \le 1/2T_b \\ 0, & \text{en otro caso} \end{cases} \tag{4.68}$$

De tal modo, la respuesta en frecuencia completa del esquema de transmisión de señales duobinarias tiene la forma de una función coseno de medio ciclo, de acuerdo con

$$H_1(f) = \begin{cases} 2\cos(\pi fT_b)\exp(-j\pi fT_b), & |f| \le 1/2T_b \\ 0, & \text{en otro caso} \end{cases} \tag{4.69}$$

para la cual la respuesta en magnitud y la respuesta en fase se muestran en la figura 4.12*a* y 4.12*b*, respectivamente. Una ventaja de esta respuesta en frecuencia es que en la práctica puede aproximarse fácilmente porque hay continuidad en los bordes de la banda.

De la primera línea de la ecuación (4.67) y de la definición de $H_{\text{Nyquist}}(f)$ en la ecuación (4.68), encontramos que la respuesta al impulso correspondiente a la respuesta en frecuencia

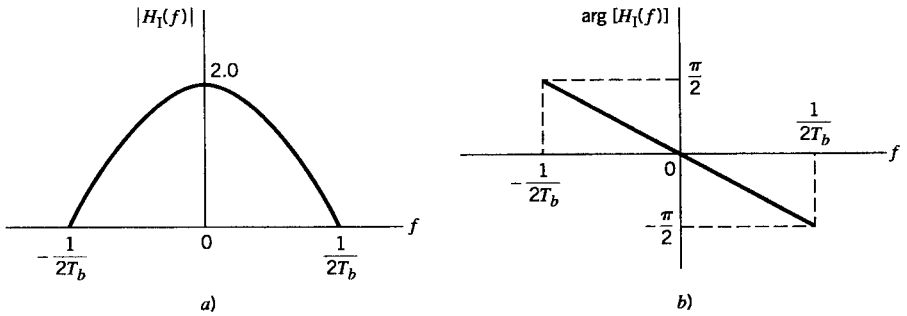


FIGURA 4.12 Respuesta en frecuencia de un filtro de conversión duobinario. *a*) Respuesta en magnitud. *b*) Respuesta en fase.

$H_I(f)$ está compuesta por dos pulsos senc (Nyquist) que están desplazados en el tiempo por T_b segundos uno con respecto al otro, como se indica mediante (salvo por el factor de escalamiento)

$$\begin{aligned} h_I(t) &= \frac{\text{sen}(\pi t/T_b)}{\pi t/T_b} + \frac{\text{sen}[\pi(t-T_b)/T_b]}{\pi(t-T_b)/T_b} \\ &= \frac{\text{sen}(\pi t/T_b)}{\pi t/T_b} - \frac{\text{sen}(\pi t/T_b)}{\pi(t-T_b)/T_b} \\ &= \frac{T_b^2 \,\text{sen}(\pi t/T_b)}{\pi t(T_b - t)} \end{aligned} \tag{4.70}$$

La respuesta al impulso $h_I(t)$ se grafica en la figura 4.13, donde vemos que ésta sólo tiene dos valores distinguibles en los instantes de muestreo. La forma de $h_I(t)$ que se muestra aquí explica por qué nos referimos también a este tipo de codificación correlativa como transmisión de señales de respuesta parcial. La respuesta a un pulso de entrada se distribuye sobre más de un intervalo de transmisión de señales; establecido de otra manera, la respuesta en cualquier intervalo de transmisión de señal es "parcial". Note también que los extremos de $h_I(t)$ decaen como $1/|t|^2$, que es una razón de decaimiento superior que la relativa a $1/|t|$ que se encuentra en el canal de Nyquist ideal.

La secuencia original de dos niveles $\{a_k\}$ puede detectarse a partir de la secuencia codificada duobinaria $\{c_k\}$ apelando al uso de la ecuación (4.66). De modo específico, dejemos que $\hat{a}_k$ represente la *estimación* del pulso original a_k como se consigue en el receptor en el tiempo $t = kt_b$. Entonces, sustrayendo la estimación previa $\hat{a}_{k-1}$ de c_k, obtenemos

$$\hat{a}_k = c_k - \hat{a}_{k-1} \tag{4.71}$$

Es claro que si c_k se recibe sin error y si también la estimación previa $\hat{a}_{k-1}$ en el tiempo $t = (k-1)T_b$ corresponde a una decisión correcta, entonces la estimación presente $\hat{a}_k$ será también

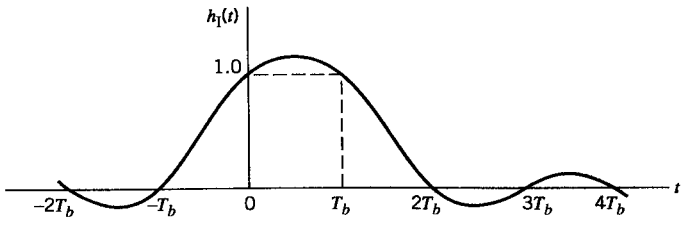


FIGURA 4.13 Respuesta al impulso de un filtro de conversión duobinario.

correcta. La técnica relativa al uso de una estimación almacenada del símbolo anterior se conoce como *realimentación de decisión*.

Observamos que el procedimiento de detección que acabamos de describir es en esencia el inverso de la operación del filtro de retardo de línea simple en el transmisor. Sin embargo, una desventaja principal de este procedimiento de detección es que una vez que se producen los errores, éstos tienden a *propagarse* por la salida debido a que la decisión sobre la entrada presente a_k depende de la corrección de la decisión que se tomó en la entrada anterior a_{k-1}.

Una manera práctica de evitar el fenómeno de la propagación de errores consiste en utilizar *precodificación* antes de la codificación duobinaria, como se indica en la figura 4.14. La operación de precodificación realizada sobre la secuencia de datos binarios $\{b_k\}$ convierte ésta en otra secuencia binaria $\{d_k\}$ definida por

$$d_k = b_k \oplus d_{k-1} \tag{4.72}$$

donde el símbolo $\oplus$ denota la *adición de módulo-dos* de los dígitos binarios b_k y d_{k-1}. Esta adición es equivalente a una operación OR EXCLUSIVA de dos entradas, que se lleva a cabo del modo siguiente:

$$d_k = \begin{cases} \text{símbolo } 1 & \text{si ya sea el símbolo } b_k \text{ o el símbolo } d_{k-1} \text{ (pero no ambos) es 1} \\ \text{símbolo } 0 & \text{en otro caso} \end{cases} \tag{4.73}$$

La secuencia binaria precodificada $\{d_k\}$ se aplica a un modulador por amplitud de pulsos, produciendo una secuencia correspondiente de dos niveles de pulsos cortos $\{a_k\}$, donde $a_k = \pm 1$ como antes. Esta secuencia de pulsos cortos se aplica después a un codificador duobinario, produciendo de esa manera la secuencia $\{c_k\}$ que se relaciona con $\{a_k\}$ como sigue

$$c_k = a_k + a_{k-1} \tag{4.74}$$

Advierta que a diferencia de la operación lineal de la codificación duobinaria, la precodificación descrita por la ecuación (4.72) es una operación *no lineal*.

El uso combinado de las ecuaciones (4.72) y (4.74) produce

$$c_k = \begin{cases} 0 & \text{si el símbolo de datos } b_k \text{ es 1} \\ \pm 2 & \text{si el símbolo de datos } b_k \text{ es 0} \end{cases} \tag{4.75}$$

la cual se ilustra en el ejemplo 4.3. A partir de la ecuación (4.75) deducimos la siguiente regla de decisión para detectar la secuencia binaria original $\{b_k\}$ a partir de $\{c_k\}$:

$$\begin{aligned} &\text{Si } |c_k| < 1, \quad \text{indica que el símbolo } b_k \text{ es 1} \\ &\text{Si } |c_k| > 1, \quad \text{indica que el símbolo } b_k \text{ es 0} \end{aligned} \tag{4.76}$$

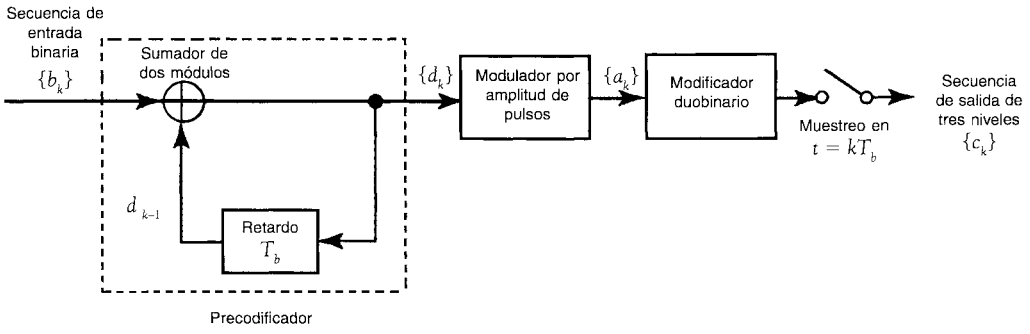


FIGURA 4.14 Un esquema duobinario precodificado; los detalles del codificador duobinario se presentan en la figura 4.11.

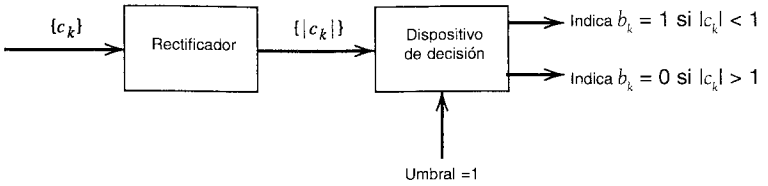


FIGURA 4.15 Detector para recuperar la secuencia binaria original a partir de la salida precodificada del codificador duobinario.

Cuando $|c_k| = 1$, el receptor simplemente realiza una suposición aleatoria en favor del símbolo 1 o 0. De acuerdo con esta regla de decisión, el detector consiste en un rectificador, cuya salida se compara en un dispositivo de decisión con respecto a un umbral de 1. En la figura 4.15 se presenta un diagrama de bloques del detector. Una característica útil del mismo es que no se requiere conocimiento de ninguna muestra de entrada aparte de la presente. Por tanto, la propagación de errores no puede ocurrir en el detector de la figura 4.15.

EJEMPLO 4.3 Codificación duobinaria con precodificación

Considere la secuencia de datos binarios 0010110. Para proceder con la precodificación de esta secuencia, que implica alimentar la salida del precodificador de vuelta a la entrada, sumamos un bit extra a la salida del precodificador. Este bit extra se elige arbitrariamente igual a 1. Por tanto, al utilizar la ecuación (4.73) encontramos que la secuencia $\{d_k\}$ a la salida del precodificador es como se muestra en el renglón 2 de la tabla 4.1. La representación polar de la secuencia precodificada $\{d_k\}$ se muestra en el renglón 3 de la tabla 4.1. Por último, utilizando la ecuación (4.74), encontramos que la salida del codificador duobinario tiene los niveles de amplitud dados en el renglón 4 de la tabla 4.1.

Para detectar la secuencia binaria original aplicamos la regla de decisión de la ecuación (4.76), y obtenemos de ese modo la secuencia binaria dada en el renglón 5 de la tabla 4.1. Este último resultado demuestra que, a falta de ruido, la secuencia binaria original se detecta de modo correcto. ◀

TRANSMISIÓN MODIFICADA DE SEÑALES DUOBINARIAS

En la técnica de transmisión de señales duobinarias la respuesta en frecuencia $H(f)$ y, en consecuencia, la densidad espectral de potencia del pulso transmitido, no son cero en el origen. Esto se considera como un rasgo indeseable en algunas aplicaciones, ya que muchos canales de comunicación no pueden transmitir una componente de cd. Es posible corregir esta deficiencia utilizando la *respuesta parcial de clase IV* o la técnica *duobinaria modificada*, la cual implica un tramo de correlación de dos dígitos binarios. Esta forma especial de correlación se consigue sustrayendo pulsos modulados en amplitud espaciados a $2T_b$ segundos, como se indica en el diagrama de bloques de la figura 4.16. El codificador incluye un retardo de $2T_b$ segundos. La

TABLA 4.1 *Ilustración del ejemplo 4.3 acerca de la codificación duobinaria*

Secuencia binaria $\{b_k\}$		0	0	1	0	1	1	0
Secuencia precodificada $\{d_k\}$	1	1	1	0	0	1	0	0
Secuencia de dos niveles $\{a_k\}$	+1	+1	+1	−1	−1	+1	−1	−1
Salida del codificador duobinario $\{c_k\}$		+2	+2	0	−2	0	0	−2
Secuencia binaria obtenida al aplicar la regla de decisión de la ecuación (4.76)		0	0	1	0	1	1	0

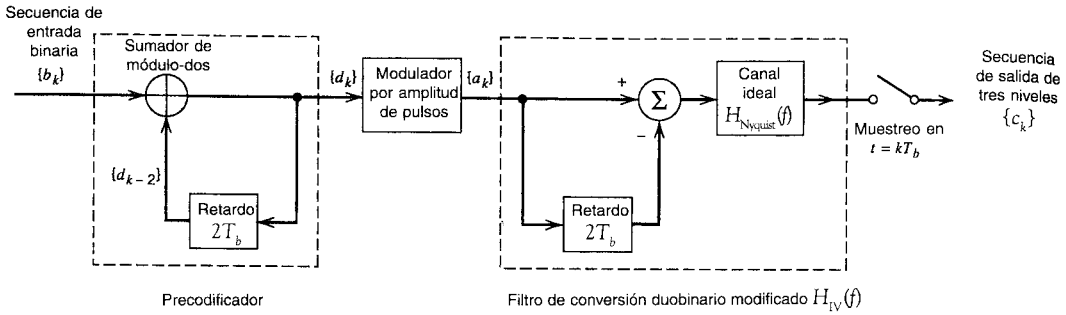


FIGURA 4.16 Esquema de transmisión de señales duobinarias modificado.

salida del filtro de conversión duobinaria modificada se relaciona con la secuencia de entrada de dos niveles $\{a_k\}$ a la salida del modulador por amplitud de pulsos de la manera siguiente:

$$c_k = a_k - a_{k-2} \tag{4.77}$$

De nuevo encontramos aquí que se genera una señal de tres niveles con $a_k = \pm 1$, vemos que c_k toma uno de tres valores: $+2$, 0 y -2.

La respuesta en frecuencia completa del filtro de retardo de línea conectado en cascada con un canal de Nyquist ideal, como en la figura 4.16, está dada por

$$\begin{aligned} H_{IV}(f) &= H_{Nyquist}(f)[1 - \exp(-j4\pi f T_b)] \\ &= 2jH_{Nyquist}(f)\,\text{sen}(2\pi f T_b)\exp(-j2\pi f T_b) \end{aligned} \tag{4.78}$$

donde el subíndice IV en $H_{IV}(f)$ indica la clase pertinente de respuesta parcial y $H_{Nyquist}(f)$ se define como la ecuación (4.68). Por tanto, tenemos una respuesta en frecuencia completa en la forma de una función seno de medio ciclo, como se indica mediante

$$H_{IV}(f) = \begin{cases} 2j\,\text{sen}(2\pi f T_b)\exp(-j2\pi f T_b), & |f| \le 1/2T_b \\ 0, & \text{en otro lado} \end{cases} \tag{4.79}$$

La respuesta en magnitud y la respuesta en fase correspondientes del codificador duobinario modificado se muestran en las figuras 4.17*a* y 4.17*b*, respectivamente. Un rasgo útil del codificador duobinario modificado es debido a que su salida no tiene componente de cd. Advierta también que esta segunda forma de codificación de nivel correlativo exhibe la misma continuidad en los bordes de la banda que en el caso de la transmisión de señales duobinaria.

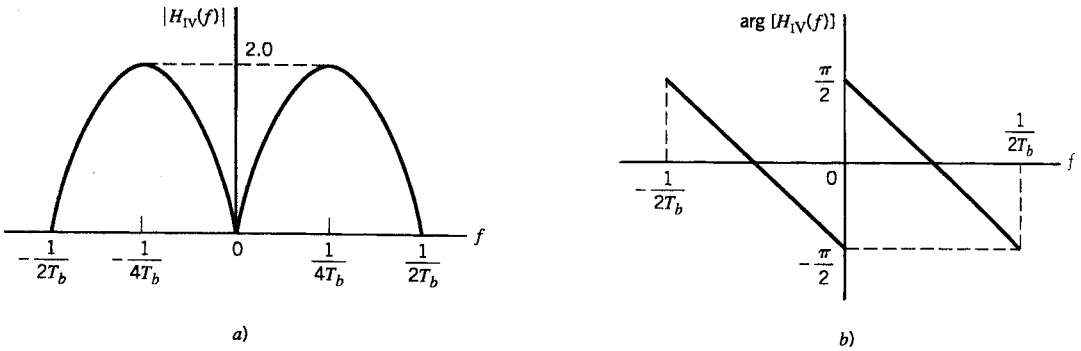


FIGURA 4.17 Respuesta en frecuencia del filtro de conversión duobinario modificado. *a*) Respuesta en magnitud. *b*) Respuesta en fase.

De la primera línea de la ecuación (4.78) y de la definición de $H_{\text{Nyquist}}(f)$ en la ecuación (4.68), encontramos que la respuesta al impulso del codificador duobinario modificado consiste en dos pulsos senc (Nyquist) que están desplazados en el tiempo por $2T_b$ segundos uno con respecto al otro (excepto por un factor de escalamiento), como se muestra por medio de

$$\begin{aligned} h_{\text{IV}}(t) &= \frac{\operatorname{sen}(\pi t/T_b)}{\pi t/T_b} - \frac{\operatorname{sen}[\pi(t-2T_b)/T_b)]}{\pi(t-2T_b)/T_b} \\ &= \frac{\operatorname{sen}(\pi t/T_b)}{\pi t/T_b} - \frac{\operatorname{sen}(\pi t/T_b)}{\pi(t-2T_b)/T_b} \\ &= \frac{2T_b^2 \operatorname{sen}(\pi t/T_b)}{\pi t(2T_b - t)} \end{aligned} \tag{4.80}$$

Esta respuesta al impulso se grafica en la figura 4.18, la cual muestra que tiene *tres* niveles distinguibles en los instantes de muestreo. Advierta también que, como con la transmisión de señales duobinaria, las colas de $h_{\text{IV}}(t)$ para la transmisión de señales duobinaria modificada decae como $1/|t|^2$.

Para eliminar la posibilidad de propagación de errores en el sistema duobinario modificado, usamos un procedimiento de precodificación similar al que se empleó en el caso duobinario. Específicamente, antes de la generación de la señal duobinaria modificada, se usa adición lógica de dos módulos en las señales separadas por $2T_b$ segundos (vea el extremo frontal de la figura 4.16), de acuerdo con

$$\begin{aligned} d_k &= b_k \oplus d_{k-2} \\ &= \begin{cases} \text{símbolo } 1 & \text{si el símbolo } b_k \text{ o el símbolo } d_{k-2} \text{ (pero no ambos) es 1} \\ \text{símbolo } 0 & \text{en otro caso} \end{cases} \end{aligned} \tag{4.81}$$

donde $\{b_k\}$ es la secuencia de datos binarios entrante y $\{d_k\}$ es la secuencia de la salida del precodificador. La secuencia precodificada $\{d_k\}$ producida de ese modo se aplica entonces a un modulador por amplitud de pulsos y luego al filtro de conversión duobinaria modificada.

En la figura 4.16, el dígito de salida c_k es igual a -2, 0 o $+2$ suponiendo que el modulador por amplitud de pulsos utiliza una representación polar para la secuencia precodificada $\{d_k\}$. Encontramos también que el dígito detectado $\hat{b}_k$ en la salida de receptor puede extraerse de c_k descartando la polaridad de c_k. De modo específico, podemos formular la siguiente regla de decisión:

$$\begin{aligned} &\text{Si } |c_k| > 1, \quad \text{indica que el símbolo } b_k \text{ es 1} \\ &\text{Si } |c_k| < 1, \quad \text{indica que el símbolo } b_k \text{ es 0} \end{aligned} \tag{4.82}$$

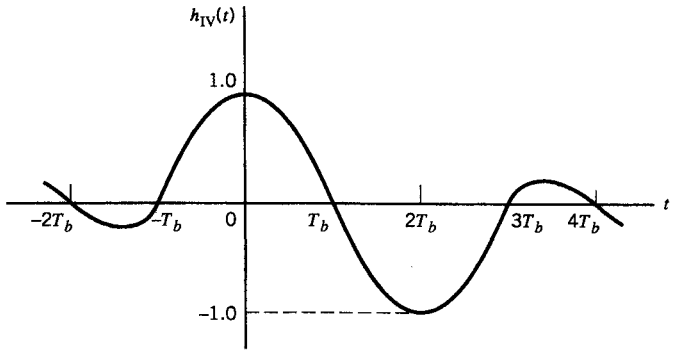


FIGURA 4.18 Respuesta de impulso del filtro de conversión duobinario modificado.

Cuando $|c_k| = 1$, el receptor realiza una suposición aleatoria en favor del símbolo 1 o 0. Como con la transmisión de señales duobinaria, es posible advertir lo siguiente:

- A falta de ruido del canal, la secuencia binaria detectada $\{\hat{b}_k\}$ es exactamente la misma que la secuencia binaria original $\{b_k\}$ a la entrada del transmisor.
- El uso de la ecuación (4.81) requiere la suma de dos bits adicionales a la secuencia precodificada $\{a_k\}$. La composición de la secuencia decodificada $\{\hat{b}_k\}$ al utilizar la ecuación (4.82) es invariante a la selección que se realiza de estos dos bits.

■ FORMA GENERALIZADA DE LA CODIFICACIÓN DE NIVEL CORRELATIVO (TRANSMISIÓN DE SEÑALES DE RESPUESTA PARCIAL)

Las técnicas duobinaria y duobinaria modificada tienen tramos de correlación de 1 dígito binario y de 2 dígitos binarios, respectivamente. De modo directo se generalizan estas dos técnicas para otros esquemas, los cuales se conocen colectivamente como *esquemas de codificación de nivel correlativo* o de *transmisión de señales de respuesta parcial*. Esta generalización se muestra en la figura 4.19, donde $H_{\text{Nyquist}}(f)$ se define en la ecuación (4.68). En ella se incluye el uso de un filtro de retardo de línea en derivación con pesos de derivación $w_0, w_1, \cdots, w_{N-1}$. De modo específico, es posible conseguir diferentes clases de esquemas de transmisión de señales de res-

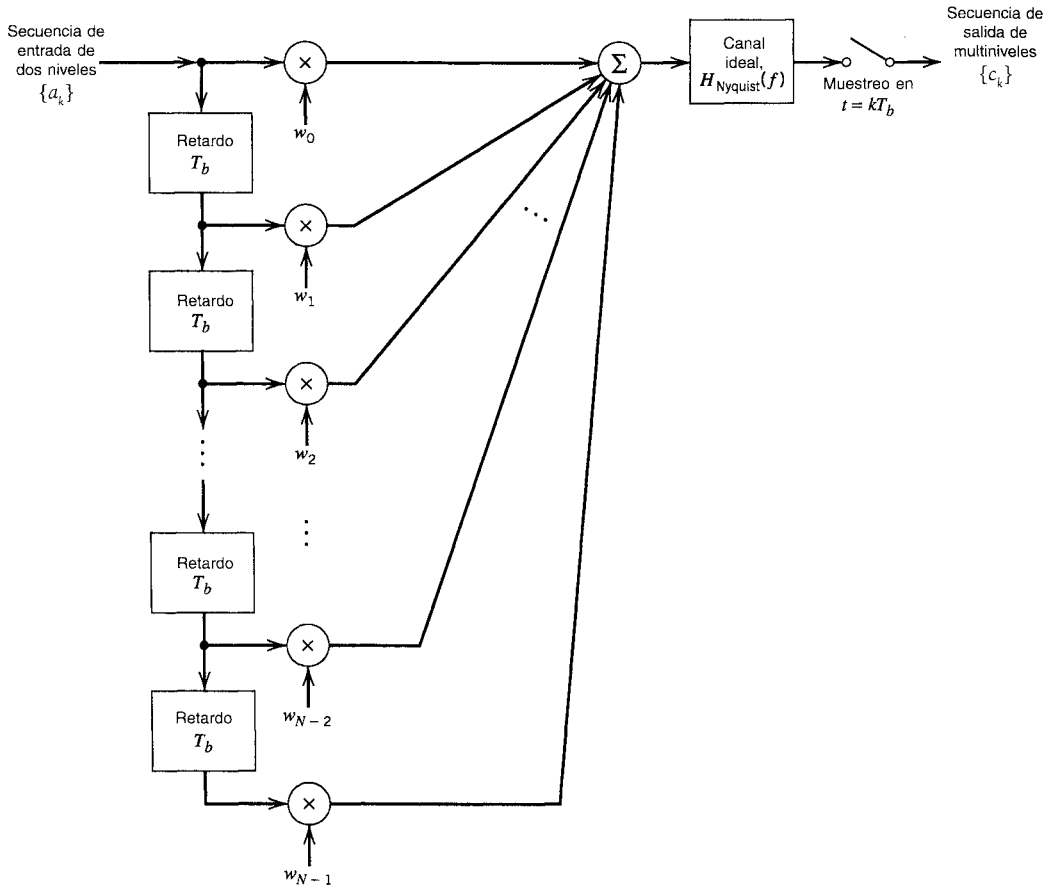


FIGURA 4.19 Esquema de codificación correlativa generalizado.

TABLA 4.2 ***Clases diferentes de esquemas de transmisión de señales de respuesta parcial en referencia a la figura 4.19***

Tipo de clase	N	w_0	w_1	w_2	w_3	w_4	*Comentarios*
I	2	1	1				Codificación duobinaria
II	3	1	2	1			
III	3	2	1	−1			
IV	3	1	0	−1			Codificación duobinaria modificada
V	5	−1	0	2	0	−1	

puesta parcial utilizando una combinación lineal ponderada de N pulsos de Nyquist ideales (senc), como se indica por medio de

$$h(t) = \sum_{n=0}^{N-1} w_n \operatorname{senc}\left(\frac{t}{T_b} - n\right) \tag{4.83}$$

Una elección apropiada de los pesos de derivación en la ecuación (4.83) produce una diversidad de formas espectrales ideadas para satisfacer aplicaciones individuales. La tabla 4.2 presenta los detalles específicos de cinco clases diferentes de esquemas de transmisión de señales de respuesta parcial. Por ejemplo, en el caso duobinario (respuesta parcial de clase I), tenemos

$$w_0 = +1$$
$$w_1 = +1$$

y $w_n = 0$ para $n \geq 2$. En el caso duobinario modificado (respuesta parcial de clase IV), tenemos

$$w_0 = +1$$
$$w_1 = 0$$
$$w_2 = -1$$

y $w_n = 0$ para $n \geq 3$.

Las características útiles de los esquemas de transmisión de señales de respuesta parcial pueden resumirse ahora del modo siguiente:

- La transmisión de datos binarios por un canal físico de banda base puede lograrse a una velocidad cercana a la de Nyquist, utilizando filtros realizables con características de corte graduales.
- Es posible producir formas espectrales diferentes, apropiadas a la aplicación respectiva.

Sin embargo, estas características deseables se consiguen a un precio: se requiere una relación señal a ruido mayor para producir la misma probabilidad promedio de error de símbolo en la presencia de ruido que en los correspondientes sistemas PAM binarios debido a un aumento en el número de niveles de señal que se utilizan.

4.7 *Transmisión PAM M-aria de banda base*

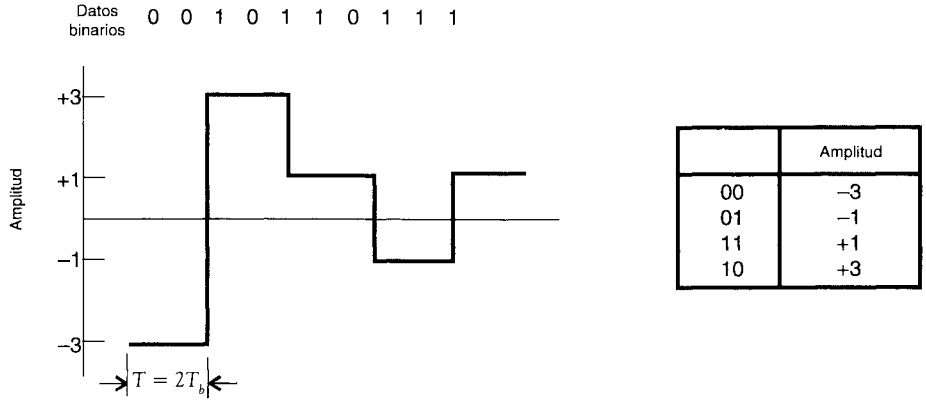
En el sistema PAM binario de banda base de la figura 4.7, el modulador por amplitud de pulsos produce pulsos binarios, esto es, pulsos con uno de dos niveles de amplitud posible. Por

otra parte, en un *sistema PAM M-ario de banda base*, el modulador por amplitud de pulsos produce uno de M niveles de amplitud posibles con M > 2. Esta fórmula de modulación por pulsos se ilustra en la figura 4.20*a* para el caso de un sistema *cuaternario* (M = 4) y una secuencia de datos binarios 0010110111. La forma de onda mostrada en la figura 4.20*a* se basa en la representación eléctrica para cada uno de los cuatro posibles *dibits* (pares de bits) dados en la figura 4.20*b*. Advierta que esta representación es de *codificación Gray*, lo que significa que cualquier dibit en el alfabeto cuaternario difiere de un dibit adyacente en una única posición de bit.

En un sistema M-ario, la fuente de información emite una secuencia de símbolos que provienen de un alfabeto compuesto por M símbolos. Cada nivel de amplitud a la salida del modulador por amplitud de pulsos corresponde a un símbolo distinto, por lo que existen M niveles de amplitud distintos que se van a transmitir. Considere el sistema PAM M-ario con un alfabeto de señales que contiene M símbolos de igual probabilidad y estadísticamente independientes, con la duración del símbolo denotada por T segundos. Nos referiremos a $1/T$ como la *velocidad de transmisión de señales* del sistema, la cual se expresa en *símbolos por segundo* o *bauds*. Es informativo relacionar la velocidad de transmisión de señales de este sistema con la de un sistema PAM binario equivalente para el cual el valor de M es 2 y los símbolos binarios sucesivos 1 y 0 son de igual probabilidad y estadísticamente independientes, con la duración de cualquier símbolo denotada por T_b segundos. Con base en la condición que se acaba de describir, el sistema PAM binario produce información a razón de $1/T_b$ bits por segundo. Observamos además que en el caso de un sistema PAM cuaternario, por ejemplo, los cuatro símbolos posibles pueden identificarse con los dibits 00, 01, 10 y 11. Vemos de este modo que cada símbolo representa 2 bits de información, y que 1 baud es igual a 2 bits por segundo. Es posible generalizar este resultado estableciendo que en un sistema PAM M–ario, un baud es igual a $\log_2 M$ bits por segundo, y la duración del símbolo T del sistema PAM M-ario se relaciona con la duración de bit T_b del sistema PAM binario equivalente en la forma

$$T = T_b \log_2 M \tag{4.84}$$

Por tanto, en un ancho de banda de canal dado, encontramos que al utilizar un sistema PAM M-ario somos capaces de transmitir información a una velocidad que es $\log_2 M$ más rápida que el sistema PAM binario correspondiente. Sin embargo, para realizar la misma probabilidad promedio del error de símbolo, un sistema PAM M-ario requiere mayor potencia transmitida. De manera específica, encontramos que para M mucho mayor que 2 y una probabilidad promedio



	Amplitud
00	−3
01	−1
11	+1
10	+3

FIGURA 4.20 Salida de un sistema cuaternario. *a*) Forma de onda. *b*) Representación de los cuatro posibles dibits, con base en la codificación Gray.

del error de símbolo pequeña comparada con 1, la potencia transmitida debe incrementarse por el factor $M^2/\log_2 M$, en comparación con el sistema PAM binario.

En un sistema M-ario de banda base, antes que todo, la secuencia de símbolos emitida por la fuente de información se convierte en un tren de pulsos PAM de nivel M por medio de un modulador por amplitud de pulsos a la entrada del transmisor. Después, como en el caso del sistema PAM binario, se da forma a este tren de pulsos mediante un filtro de transmisión y luego se transmite por el canal de comunicación, lo cual corrompe la forma de onda de la señal tanto con ruido como con distorsión. La señal recibida se hace pasar por un filtro de recepción y luego se muestrea a una velocidad apropiada de sincronización con el transmisor. Cada muestra se compara con los valores de *umbral* preestablecidos (denominados también niveles de *división*), y se efectúa una decisión relativa a cuál símbolo se transmitió. Por tanto, encontramos que los diseños del modulador por amplitud de pulsos y el dispositivo de toma de decisiones en una PAM M-aria resultan más complejos que los correspondientes al sistema PAM binario. La interferencia de intersímbolo, el ruido y la sincronización imperfecta provocan la aparición de errores en la salida del receptor. Los filtros de transmisión y recepción se diseñan para minimizar estos errores. Los procedimientos que se utilizan para diseñar estos filtros son similares a los que se señalaron en las secciones 4.5 y 4.6 para los sistemas PAM binarios de banda base.

4.8 *Líneas digitales de suscriptor*

En este punto de la transmisión de datos de banda base es muy apropiado que nos apartemos de los aspectos teóricos del estudio y consideremos una aplicación que crece rápidamente: *las líneas digitales de suscriptor*.[6] Una línea digital de suscriptor (DSL) opera por un lazo local (menor que 1.5 km) que proporciona una conexión directa entre una terminal de usuario (por ejemplo, una computadora) y una *oficina central* de la compañía telefónica, como se ilustra en la figura 4.21. A través de dicha oficina, un usuario de DSL se conecta a la red de datos troncal de banda ancha, la cual se basa en tecnologías como el modo de transferencia asíncrona (ATM) y el protocolo de Internet (IP); estas tecnologías y los recursos de red relacionados (es decir, fibras ópticas, SONET) se explican en el capítulo de antecedentes y preliminares. Por tanto, la señal que contiene la información se mantiene en el dominio digital en todo momento desde la terminal de usuario hasta el proveedor del servicio de Internet, conmutándose o direccionándose a intervalos regulares en el transcurso de la transmisión a través de la red de datos.

En apego al interés de una puesta en práctica económica, las líneas digitales de suscriptor utilizan pares trenzados configurados para ofrecer una elevada velocidad de datos, dúplex total, capacidad de transmisión digital. (Los pares trenzados se utilizan también en las comunicaciones telefónicas ordinarias, como se explica en el capítulo de antecedentes y preliminares.) Para conseguir la transmisión de dos alambres en dúplex completa, podemos utilizar dos modos posibles de operación:

1. *Multiplexado por compresión de tiempo (TC)*, donde la transmisión de datos en las dos direcciones opuestas sobre la línea común se separa en el tiempo. De modo específico, se

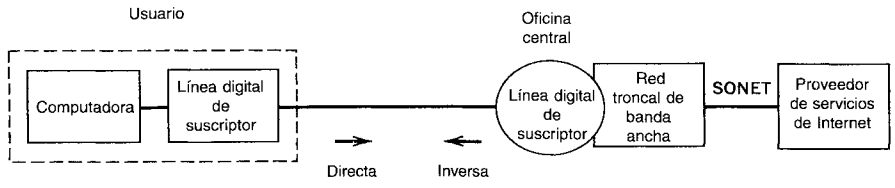


FIGURA 4.21 Diagrama de bloques que describe el ambiente operativo de las líneas digitales de suscriptor.

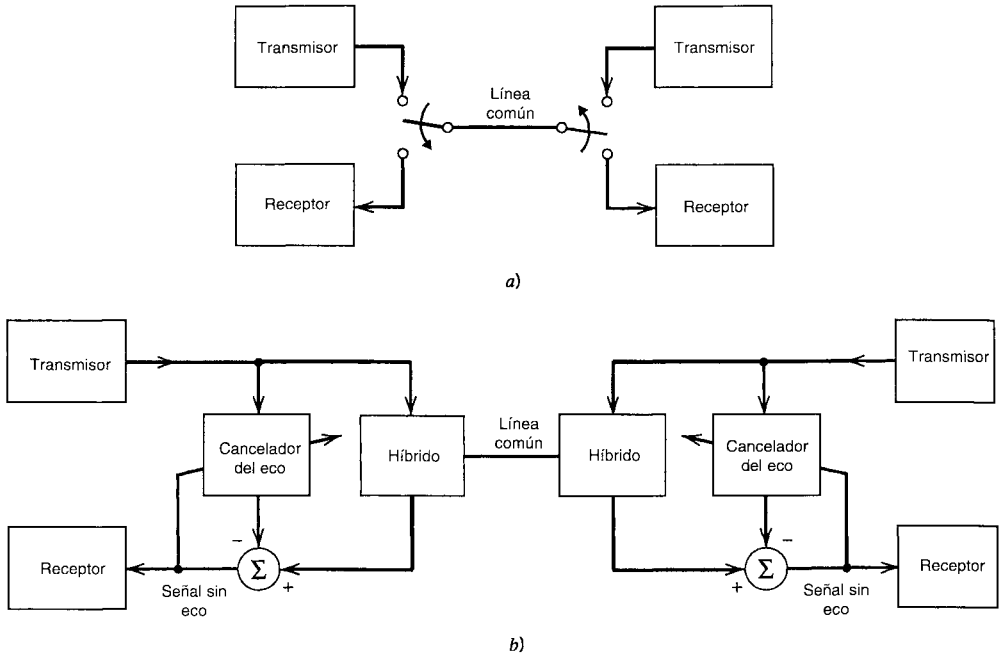


FIGURA 4.22 Operación dúplex completa utilizando *a*) multiplexado de compresión de tiempo y *b*) cancelación del eco.

envían en forma alternada bloques de datos en ráfagas en cada dirección, como se ilustra en la figura 4.22*a*. Para tener en cuenta el tipo de propagación a través de la línea, se inserta un tiempo de seguridad entre las ráfagas de datos individuales. En consecuencia, la velocidad de la línea es un poco mayor que el doble de la velocidad de los datos.

2. *Modo de cancelación del eco*, el cual soporta el flujo simultáneo de datos a lo largo de la línea común en ambas direcciones. Para que esta forma de transmisión sea factible, cada *transceptor* (transmisor/receptor) incluye un híbrido para dos propósitos: la separación de la señal transmitida de la señal recibida y la conversión de dos a cuatro alambres, como se muestra en la figura 4.22*b*. El *híbrido*, o más precisamente el *transformador híbrido*, es básicamente un circuito puente con tres puertos (pares de terminales), como se presenta en la figura 4.23. Si el puente no está perfectamente balanceado, el puerto transmisor del híbri-

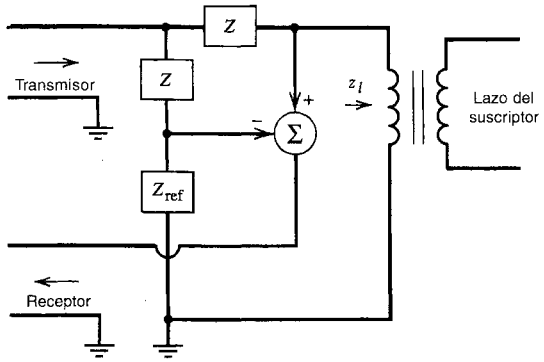


FIGURA 4.23 Circuito simplificado de un transformador híbrido. Para que el puente esté balanceado, la impedancia de referencia Z_{ref} debe ser igual a la impedancia de línea z_l.

do se acopla al puerto receptor, originando por ello un *eco* que surge de la fuga de la señal transmitida del extremo cercano (local) hacia el receptor del extremo cercano (local). Para cancelar el eco indeseable, cada transceptor incluye un *cancelador del eco*, como se muestra en la figura 4.22*b*. Puesto que los datos pueden fluir a través de la línea de manera simultánea en varias direcciones, la velocidad de línea es la misma que la de los datos.

De acuerdo con lo expuesto, es claro que los esquemas de cancelación del eco ofrecen un mucho mejor desempeño de transmisión de datos que el esquema de multiplexado por compresión de tiempo, aunque a costa de una mayor complejidad. Sin embargo, al implementar el transceptor completo en un solo chip a muy grande escala de integración (VLSI), el costo puede afrontarse a pesar de la mayor complejidad. En Estados Unidos, el esquema de cancelación del eco ha sido adoptado con base en el diseño de transceptores. Una puesta en práctica adaptable del cancelador de eco se explica en el problema 4.31.

Además del eco, existen otros deterioros del medio de transmisión que necesitan considerarse. Los dos dominantes son la interferencia de intersímbolo y la diafonía, los cuales se explican a continuación.

Hasta una aproximación de primer orden, la respuesta en magnitud al cuadrado de un par trenzado está dada por

$$|H(f)|^2 = \exp(-\alpha\sqrt{f}) \tag{4.85}$$

donde

$$\alpha = k\frac{l}{l_0} \tag{4.86}$$

En la ecuación (4.85), la frecuencia f se mide en kHz, k es una constante física del par trenzado, l_0 es la longitud de referencia (por ejemplo, kilómetros) y l es la longitud real del par trenzado. La ecuación (4.85) señala un mayor deterioro en el uso de un par trenzado para la transmisión de datos de banda base: la caída gradual en la respuesta en frecuencia, la cual, a su vez, origina la interferencia de intersímbolo.

Con respecto a la diafonía, su causa primaria de ocurrencia es el acoplamiento capacitivo que existe entre pares trenzados adyacentes en un cable. Por lo común, de cinco a siete de los pares trenzados más cercanos en el cable provocan la mayor parte de la diafonía. En cualquier caso, se observan dos tipos de este fenómeno en un receptor de interés:

1. *Diafonía del extremo cercano (NEXT)*; la generan los transmisores ubicados en el mismo extremo del cable que el receptor, como se ilustra en la figura 4.24*a*.

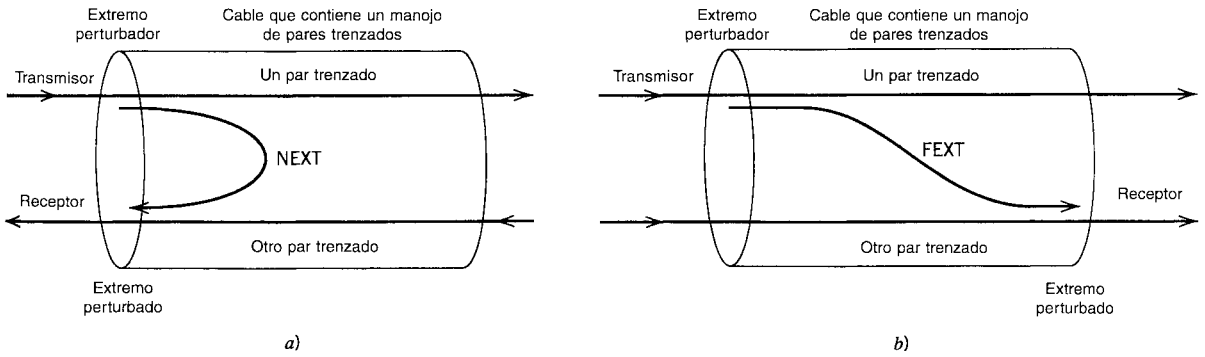


FIGURA 4.24 *a*) Diafonía del extremo cercano (NEXT). *b*) Diafonía del extremo lejano (FEXT).

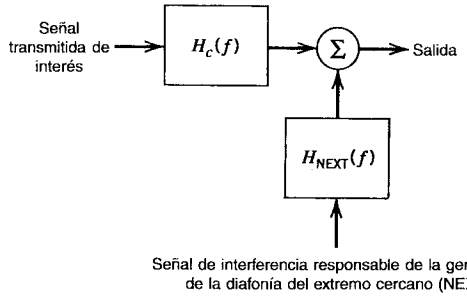


FIGURA 4.25 Modelo del canal de par trenzado.

2. *Diafonía del extremo lejano (FEXT)*, que resulta de los transmisores ubicados lejos del receptor, como se indica en la figura 4.24*b*.

La FEXT se somete de forma natural a las mismas pérdidas de línea que la señal, aunque eso no ocurre con la NEXT. Por tanto, en el esquema de cancelación del eco de la figura 4.22*b*, donde las señales viajan en ambas direcciones en el cable, la NEXT será mucho más intensa que la FEXT. De aquí en adelante ignoraremos el efecto de la FEXT.

En realidad, la diafonía del extremo cercano y la interferencia de intersímbolo son dos de los factores más importantes en la determinación del desempeño de un lazo digital de suscriptor. La figura 4.25 presenta el modelo de un canal de par trenzado dominado por estos dos deterioros. Puesto que todos los pares trenzados suelen transmitir señales similares, podemos modelar la NEXT como una señal de la misma densidad espectral de potencia que la señal transmitida que atraviesa una respuesta en frecuencia *de diafonía* $H_{\text{NEXT}}(f)$, la cual se aproxima por medio de

$$H_{\text{NEXT}}(f) = \beta f^{3/2} \tag{4.87}$$

donde β es una constante del cable. El punto interesante que advertir de la figura 4.25 es que tanto la señal transmitida como la de interferencia tienen la misma densidad espectral de potencia; difieren entre sí únicamente de sus respuestas en frecuencia asociadas, como indican las ecuaciones (4.85) y (4.87), respectivamente. Cuando el modelo que se describió aquí se utiliza en estudios de simulación, la señal transmitida se representa por medio de una secuencia de datos aleatorios, y la interferencia por medio de una secuencia de ruido gaussiano.

■ CÓDIGOS DE LÍNEA PARA LÍNEAS DIGITALES DE SUSCRIPTOR

Luego de que hemos identificado los principales deterioros de transmisión, es posible describir las características deseables que debe exhibir el espectro de una señal transmitida:

1. La densidad espectral de potencia de la señal transmitida debe ser cero a la frecuencia cero, puesto que no es posible la transmisión de cd a través de un transformador híbrido.
2. La densidad espectral de potencia de la señal transmitida debe ser baja a frecuencias elevadas, por las siguientes razones:
   - La atenuación de la transmisión en un par trenzado es más severa a frecuencias elevadas.
   - La diafonía entre pares trenzados adyacentes aumenta en forma considerable a frecuencias altas debido al creciente acoplamiento capacitivo. A este respecto, recuerde que la impedancia de un capacitor es inversamente proporcional a la frecuencia.

   Para satisfacer estas propiedades deseables, debemos tener cuidado al elegir el *código de línea*, que establece la correspondencia entre la secuencia entrante de bits de los datos y los pulsos eléctricos para la transmisión en la línea. Existen varias posibilidades, cada una

con sus propias ventajas y desventajas, para una elección de este tipo. La lista de candidatos potenciales para los códigos de línea incluye los siguientes:

- *Código Manchester*, que es simple y tiene componente de cd cero. Su desventaja es la ocupación de un gran espectro, lo que hace que sea vulnerable ante la diafonía de extremo cercano y la interferencia de intersímbolo. (El código Manchester se explicó en la sección 3.7.)
- *Código duobinario modificado*, que tiene cd cero, es eficiente en forma moderada en lo que se refiere al espectro y provoca mínima interferencia de intersímbolo. Sin embargo, los estudios de simulación del desempeño relativo a la diafonía del código duobinario modificado han mostrado que su inmunidad a la diafonía de extremo cercano y a la interferencia de intersímbolo es casi de 2 o 3 dB más pobre que los códigos de bloques en los peores casos de líneas digitales de suscriptor. (El código duobinario modificado se explicó en la sección 4.6.)
- *Código bipolar*, en el cual los unos sucesivos se representan de manera alternativa por medio de niveles positivos y negativos aunque iguales, y el símbolo 0 se representa mediante el nivel cero. La transmisión de señales bipolar tiene cd cero. Las simulaciones de computadora han mostrado que el desempeño con respecto a la diafonía del extremo cercano y la interferencia de intersímbolo es un poco inferior para el código duobinario modificado en todos los lazos digitales de suscriptor. (El código bipolar, conocido también como los *códigos de inversión de marca alternada (AMI)*, se expuso en la sección 3.7.)
- *Código 2B1Q*, que significa dos dígitos binarios codificados en un símbolo cuaternario. Este código es un código de bloques que representa una señal PAM de cuatro niveles, como se ilustra en la figura 4.20. Suponiendo que los símbolos 1 y 0 son de igual probabilidad, el código 2B1Q tiene en promedio cd cero. Además, entre todos los códigos de línea considerados aquí, éste ofrece la reducción de bauds más grande, así como el mejor desempeño con respecto a la diafonía de extremo cercano y la interferencia de intersímbolo.

Debido a las propiedades deseables del código 2B1Q en comparación con el código Manchester, el código duobinario modificado y el código bipolar, así como de otros códigos de línea que no se mencionan aquí,[7] dicho código es el que se ha adoptado como el estándar en Estados Unidos para los lazos digitales de suscriptor.

Utilizando el 2B1Q como el código de línea y la puesta en práctica VLSI de un transceptor que incorpore ecualizadores adaptables y canceladores de eco, es posible conseguir una tasa de error de bits de 10^{-7} operando en dúplex completo a 160 kb/s sobre la mayoría de líneas de par trenzado de suscriptor. Una tasa de error de bits de 10^{-7} con 12 dB de margen de ruido, cuando está presente 1% de la NEXT del peor caso, es un criterio de desempeño aceptado para las líneas digitales de suscriptor. El *margen de ruido* es la cantidad de ruido del receptor (incluido el eco no cancelado) que puede tolerarse sin exceder la tasa de error de 10^{-7}.

Líneas digitales asimétricas de suscriptor

Otro tipo importante de línea digital de suscriptor es la *asimétrica* (ADSL), que es un sistema de transmisión local diseñado para soportar en forma simultánea tres servicios sobre un solo par trenzado de alambre:

1. Transmisión de *datos en sentido directo* (hacia el suscriptor) a tasas de bits de hasta 9 Mb/s.
2. Transmisión de *datos en sentido inverso* (alejándose del suscriptor) a tasas de bits de hasta 1 Mb/s.
3. Servicio telefónico convencional (POTS).

Las tasas de bits directa e inversa dependen de la longitud del par trenzado que se utiliza para realizar la transmisión. Se dice que la DSL será "asimétrica" debido a que la tasa de bits directa es muy superior a la tasa de bits inversa. La voz analógica se transmite a las frecuencias de la banda base y se combina con las transmisiones pasobanda de los datos directos e inversos utilizando

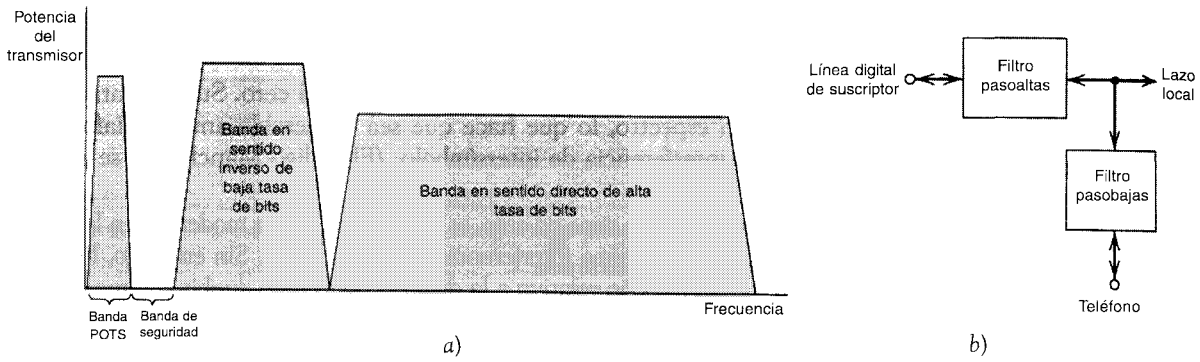


FIGURA 4.26 *a*) Ilustración de las diferentes asignaciones de banda para un sistema ADSL basado en FDM. *b*) Diagrama de bloques de divisor efectuando la función de un multiplexor o demultiplexor. Nota: ambos filtros en el divisor son bidireccionales.

multiplexado por división de frecuencia (FDM). Como se ilustra en la figura 4.26*a*, la transmisión de datos inversa se establece en una banda de frecuencia diferente de la correspondiente a la transmisión de datos directa para evitar diafonía. Además, se inserta una banda de seguridad entre la banda del servicio telefónico convencional y la banda de transmisión inversa. La coexistencia de las señales ADSL y POTS en el lazo local es posible a través del uso de pares de *divisores*; un divisor se pone en el extremo de la oficina central del lazo local y el otro en el extremo del usuario. En términos funcionales, un divisor se divide en dos filtros *bidireccionales*, como se ilustra en la figura 4.26*b*:

- Un filtro pasobajas para la transmisión de la banda base o la extracción de señales de voz.
- Un filtro pasoaltas para la transmisión pasobanda o la extracción de los datos ADSL.

En efecto, el divisor lleva a cabo la función de un multiplexor o demultiplexor por división de frecuencia, dependiendo de la dirección de transmisión de las señales.

La motivación para hacer asimétrica la DSL es ajustarse al "video sobre demanda". En tales aplicaciones, un suscriptor necesita un canal de alto rendimiento para bajar datos de video de ancho de banda grande sobre demanda desde la oficina central. En la dirección inversa, un canal de rendimiento mucho menor resulta adecuado para enviar órdenes de información, así como comandos de control en tiempo real. Por ejemplo, un ADSL para Internet que ofreciera transmisión directa a la tasa DS1 de 1.544 Mb/s y transmisión inversa de casi 160 kb/s cubriría los requerimientos de esta aplicación. La *proporción asimétrica* aproximada de 10:1 que realiza un sistema de este tipo evita que el flujo de paquetes de reconocimiento en el IP se vuelva un cuello de botella para la dirección más rápida de la transmisión de datos.

Es muy difícil transmitir datos sobre un par trenzado a una tasa DS1 y a tasas superiores, y al hacerlo de esa manera se requiere el uso de técnicas de modulación complejas. El tratamiento de este tema se deja para el capítulo 6.

4.9 *Receptor lineal óptimo*

Al resumir nuestro estudio de los sistemas de transmisión de datos de banda base que se describen en la figura 4.7, hemos tratado hasta ahora las siguientes dos condiciones del canal en forma separada:

- El *ruido del canal actuando solo*, lo cual lleva a la formulación del receptor de filtro de acoplamiento.
- La *interferencia de intersímbolo actuando sola*, que conduce a la formulación del filtro de transmisión de formación de pulsos de manera que se produzca el canal de Nyquist.

Sin embargo, en una situación real el ruido del canal y la interferencia de intersímbolo actúan juntos, afectando el comportamiento del sistema de transmisión de datos de manera combinada. En esta sección formulamos las bases para el diseño de un receptor lineal optimizado para el caso general de un canal lineal disperso y ruidoso.

En un método para diseñar un receptor lineal, el receptor se considera como un *ecualizador de forzamiento cero* seguido por un dispositivo de toma de decisiones. El objetivo de esta forma de ecualización es tener la "interferencia de intersímbolo forzada a cero" en todos los instantes $t = kT$ en los cuales se muestrea la salida del canal, salvo para $k = 0$, donde se supone que ocurre el símbolo de interés. De acuerdo con esta condición, se asegura que la detección de símbolo a símbolo será óptima en concordancia con el criterio de Nyquist, siempre que el ruido del canal $w(t)$ sea cero.

El ecualizador de forzamiento cero es relativamente fácil de diseñar debido a que ignora el efecto del ruido del canal $w(t)$. Sin embargo, una consecuencia seria de esta omisión es que la misma conduce a una degradación total del desempeño que resulta del *incremento del ruido*, fenómeno que es un rasgo inherente de la ecualización de forzamiento cero; vea el problema 4.32. Un procedimiento más refinado para el diseño del receptor corresponde a la utilización del criterio del *error cuadrático medio*, el cual ofrece una solución equilibrada al problema de reducir los efectos tanto del ruido del canal como de la interferencia de intersímbolo. En realidad, para una complejidad de cómputo preestablecida, un ecualizador diseñado sobre esta última base siempre se desempeña bien, y a menudo mucho mejor, que su contraparte de forzamiento cero. De aquí en adelante nos concentraremos en el criterio del error cuadrático medio para el diseño del receptor.

Refiriéndonos de nuevo al sistema de transmisión de datos binarios de banda base de la figura 4.7, el filtro del receptor caracterizado por la respuesta al impulso $c(t)$ produce la siguiente respuesta debida a la salida del canal $x(t)$:

$$y(t) = \int_{-\infty}^{\infty} c(\tau)x(t-\tau)\, d\tau \tag{4.88}$$

La propia salida del canal $x(t)$ está definida por

$$x(t) = \sum_{k} a_k q(t - kT_b) + w(t) \tag{4.89}$$

donde a_k es el símbolo transmitido en el tiempo $t = kT_b$ y $w(t)$ es el ruido del canal. La función del tiempo $q(t)$ es la convolución de dos respuestas al impulso: $g(t)$ pertenecientes al filtro transmisor de la formación del pulso y $h(t)$, al canal. Al sustituir la ecuación (4.89) en la (4.88) y muestrear la salida resultante $y(t)$ en el tiempo $t = iT_b$, es posible escribir

$$y(iT_b) = \xi_i + n_i \tag{4.90}$$

donde ξ_i es la componente de la señal definida por

$$\xi_i = \sum_{k} a_k \int_{-\infty}^{\infty} c(\tau)q(iT_b - kT_b - \tau)\, d\tau \tag{4.91}$$

y n_i es la componente de ruido definida por

$$n_i = \int_{-\infty}^{\infty} c(\tau)w(iT_b - \tau)\, d\tau \tag{4.92}$$

La condición para la operación perfecta del receptor radica en tener $y(iT_b) = a_i$, donde a_i es el símbolo transmitido. La desviación con respecto a esta condición produce la señal de error

$$\begin{aligned} e_i &= y(iT_b) - a_i \\ &= \xi_i + n_i - a_i \end{aligned} \tag{4.93}$$

Por tanto, podemos definir formalmente el error cuadrático medio como

$$J = \frac{1}{2}E[e_i^2] \tag{4.94}$$

donde E es el operador de la esperanza estadística, y el factor 1/2 se introduce por conveniencia en la presentación. Al sustituir la ecuación (4.93) en la (4.94) y desarrollar después los términos, obtenemos

$$J = \frac{1}{2}E[\xi_i^2] + \frac{1}{2}E[n_i^2] + \frac{1}{2}E[a_i^2] + E[\xi_i n_i] - E[n_i a_i] - E[\xi_i a_i] \tag{4.95}$$

A continuación evaluamos los seis términos de esperanza en esta ecuación en el orden que aparecen:

1. En un ambiente *estacionario*, el término cuadrático medio $E[\xi_i^2]$ es independiente del instante $t = iT_b$ al cual se muestrea la salida del filtro de recepción. Por tanto, podremos simplificar la expresión de este término escribiendo

$$E[\xi_i^2] = \sum_l \sum_k E[a_l a_k] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\tau_1)c(\tau_2)q(lT_b - \tau_1)q(kT_b - \tau_2)\, d\tau_1\, d\tau_2$$

Suponiendo que, primero, los símbolos binarios $a_k = \pm 1$ como en la ecuación (4.42) y, segundo, que los símbolos transmitidos son independientes estadísticamente, esto es,

$$E[a_l a_k] = \begin{cases} 1 & \text{para } l = k \\ 0 & \text{en otro caso} \end{cases} \tag{4.96}$$

es posible reducir aún más la expresión para el término cuadrático medio $E[\xi_i^2]$ en

$$E[\xi_i^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_q(\tau_1, \tau_2)c(\tau_1)c(\tau_2)\, d\tau_1\, d\tau_2 \tag{4.97}$$

donde

$$R_q(\tau_1, \tau_2) = \sum_k q(kT_b - \tau_1)q(kT_b - \tau_2) \tag{4.98}$$

El factor $R_q(\tau_1, \tau_2)$ es la *función de autocorrelación temporal* de la secuencia $\{q(kT_b)\}$. La calidad estacionaria de esta secuencia significa que (vea la sección 1.5)

$$R_q(\tau_1, \tau_2) = R_q(\tau_2 - \tau_1) = R_q(\tau_1 - \tau_2)$$

2. El término cuadrático medio $E[n_i^2]$ debido al ruido del canal está dado por (utilizando la ecuación (4.92))

$$\begin{aligned} E[n_i^2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\tau_1)c(\tau_2)E[w(iT_b - \tau_1)w(iT_b - \tau_2)]\, d\tau_1\, d\tau_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\tau_1)c(\tau_2)R_W(\tau_2 - \tau_1)\, d\tau_1\, d\tau_2 \end{aligned} \tag{4.99}$$

donde $R_W(\tau_2 - \tau_1)$ es la *función de autocorrelación promediada del ensamble* del ruido del canal $w(t)$. Suponiendo $w(t)$ como blanco y con densidad espectral de potencia igual a $N_0/2$, tenemos

$$R_W(\tau_2 - \tau_1) = \frac{N_0}{2}\delta(\tau_2 - \tau_1) \tag{4.100}$$

Por tanto, la expresión para $E[n_i^2]$ se simplifica en

$$E[n_i^2] = \frac{N_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\tau_1)c(\tau_2)\, \delta(\tau_2 - \tau_1)\, d\tau_1\, d\tau_2 \tag{4.101}$$

3. El término cuadrático medio $E[a_i^2]$ debido al símbolo transmitido a_i es la unidad por virtud de la ecuación (4.96); esto es,

$$E[a_i^2] = 1 \quad \text{para toda } i \tag{4.102}$$

4. La esperanza del término del producto cruz $\xi_i n_i$ es cero por dos razones: primero ξ_i y n_i son independientes y, segundo, el ruido de canal y, por tanto, n_i, tiene media cero; esto es,

$$E[\xi_i n_i] = 0 \quad \text{para toda } i \tag{4.103}$$

5. Por razones similares, la esperanza del término del producto cruz $n_i a_i$ también es cero; esto es,

$$E[n_i a_i] = 0 \quad \text{para toda } i \tag{4.104}$$

6. Por último, la esperanza del término del producto cruz $\xi_i a_i$ está dada por (utilizando la ecuación (4.91))

$$E[\xi_i a_i] = \sum_k E[a_k a_i] \int_{-\infty}^{\infty} c(\tau) q(iT_b - kT_b - \tau)\, d\tau \tag{4.105}$$

En virtud de la independencia estadística de los símbolos transmitidos que se describen en la ecuación (4.96), esta esperanza se reduce a

$$E[\xi_i a_i] = \int_{-\infty}^{\infty} c(\tau) q(-\tau)\, d\tau \tag{4.106}$$

De ese modo, sustituyendo las ecuaciones (4.97) y (4.101) en la (4.104), y (4.106) en la (4.95), es posible expresar el error medio cuadrático J por el sistema de transmisión de datos binario de la figura 4.7 como

$$J = \frac{1}{2} + \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(R_q(t-\tau) + \frac{N_0}{2}\delta(t-\tau) \right) c(t)c(\tau)\, dt\, d\tau - \int_{-\infty}^{\infty} c(t)q(-t)\, dt \tag{4.107}$$

Por conveniencia de la presentación, hemos efectuado los siguientes cambios en las variables: τ_1 y τ_2 en la primera integral se sustituyen por t y τ, respectivamente, y τ se reemplaza por t en la segunda integral. Advierta también que esta expresión para el error cuadrático medio J está realmente normalizada con respecto a la varianza de los símbolos transmitidos a_k en virtud de la suposición que se hizo en la ecuación (4.96).

Con la fórmula de la ecuación (4.107) para el error medio cuadrático J presente, estamos listos para especificar el diseño del filtro de recepción de la figura 4.7. Al diferenciar la ecuación (4.107) con respecto a la respuesta al impulso $c(t)$ del filtro de recepción, e igualar después el resultado a cero, obtenemos

$$\int_{-\infty}^{\infty} \left(R_q(t-\tau) + \frac{N_0}{2}\delta(t-\tau) \right) c(\tau)\, d\tau = q(-t) \tag{4.108}$$

La ecuación (4.108) es la fórmula para determinar la respuesta al impulso $c(t)$ del ecualizador optimizado en el sentido del error cuadrático medio. Un ecualizador diseñado de esta manera se conoce como *ecualizador de error cuadrático medio mínimo (mmse)*.

Aplicando la transformada de Fourier de ambos miembros de la ecuación (4.108), obtenemos

$$\left(S_q(f) + \frac{N_0}{2} \right) C(f) = Q^*(f) \tag{4.109}$$

donde $c(t) \rightleftharpoons C(f)$, $q(t) \rightleftharpoons Q(f)$ y $R_q \rightleftharpoons S_q(f)$. Resolviendo la ecuación (4.109) para $C(f)$, obtenemos

$$C(f) = \frac{Q^*(f)}{S_q(f) + \frac{N_0}{2}} \tag{4.110}$$

En el problema 4.33 se muestra que la densidad espectral de potencia de la secuencia $\{q(kT_b)\}$ puede expresarse como

$$S_q(f) = \frac{1}{T_b} \sum_k \left| Q\left(f + \frac{k}{T_b} \right) \right|^2 \tag{4.111}$$

lo que significa que la respuesta en frecuencia $C(f)$ del receptor lineal óptimo es *periódica*, con periodo $1/T_b$. La ecuación (4.110) sugiere la interpretación del receptor lineal óptimo como la conexión en cascada de dos componentes básicos:[8]

- Un *filtro de acoplamiento* cuya respuesta al impulso es $q(-t)$, donde $q(t) = g(t) \bigstar h(t)$.
- Un *ecualizador transversal* (retardo de línea en derivación) cuya respuesta en frecuencia es el inverso de la función periódica $S_q(f) + (N_0/2)$.

Para poner en práctica la ecuación (4.110) necesitamos exactamente un ecualizador de longitud infinita. En la práctica es posible aproximar la solución óptima utilizando un ecualizador con un conjunto finito de coeficientes $\{c_k\}_{k=-N}^{N}$, siempre que N sea lo suficientemente grande. De ese modo, el receptor toma la forma que se muestra en la figura 4.27. Advierta que el bloque marcado z^{-1} en la figura 4.27 introduce un retardo igual a T_b, que significa que el espaciamiento de la derivación del ecualizador es exactamente la misma que la duración de bit T_b. Un ecualizador configurado de este modo se dice que está en *sincronía* con el transmisor.

CONSIDERACIONES PRÁCTICAS

El receptor mmse de la figura 4.27 funciona bien en el laboratorio, donde tenemos acceso al sistema que se va a ecualizar, en cuyo caso es posible determinar un ecualizador transversal caracterizado por un conjunto de coeficientes $\{c_k\}_{k=-N}^{N}$, que proporciona una aproximación adecuada a la respuesta en frecuencia $C(f)$ de la ecuación (4.110). En un ambiente de telecomunicaciones real, el canal suele variar con el tiempo. Por ejemplo, en una red telefónica conmutada

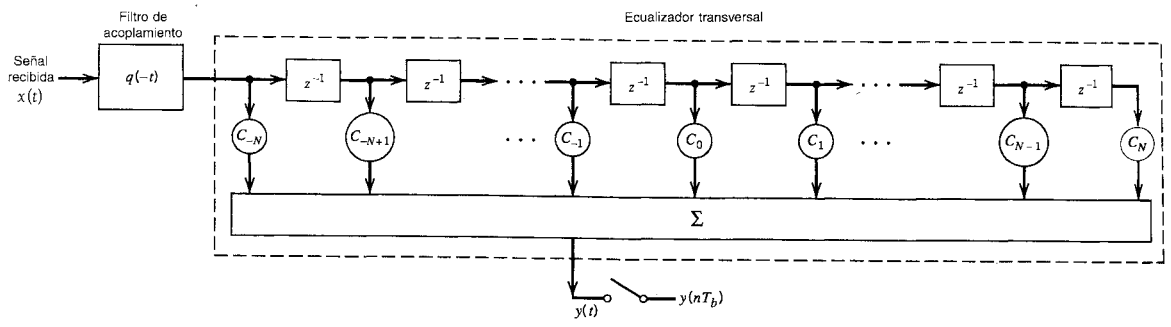


FIGURA 4.27 Receptor lineal óptimo consistente en la conexión en cascada de filtro de acoplamiento y de ecualizador transversal.

encontramos que dos factores contribuyen a la distribución de la distorsión del pulso sobre diferentes conexiones de enlace:

- Diferencias en las características de transmisión de los enlaces individuales que pueden conmutarse juntos.
- Diferencias en el número de enlaces en una conexión.

El resultado es que el canal telefónico es aleatorio porque constituye un ensamble de posibles realizaciones físicas. En consecuencia, el uso de un par fijo de filtro de acoplamiento y de ecualizador diseñados con base en las características promedio del canal tal vez no reduzca de manera adecuada los efectos de la interferencia de intersímbolo y del ruido del canal. Para conseguir la capacidad de transmisión total del canal telefónico, necesitamos un *receptor adaptable*[9] que proporcione la *puesta en práctica adaptable tanto del filtro de acoplamiento como del ecualizador de manera combinada*. El receptor es adaptable porque los coeficientes del ecualizador se ajustan de modo automático de acuerdo con un algoritmo integrado.

Otro punto de interés es que tal vez sea deseable tener las derivaciones del ecualizador espaciadas a una menor distancia que el periodo del símbolo; por lo común, el espaciamiento entre derivaciones adyacentes se hace igual a $T/2$. La estructura que resulta se conoce como un *ecualizador espaciado fraccionalmente* (*FSE*). Un FSE tiene la capacidad de compensar la distorsión del retardo de manera mucho más efectiva que un ecualizador síncrono convencional. Otra ventaja del FSE consiste en que la transmisión de los datos puede empezar con una fase de muestreo arbitraria. Sin embargo, el análisis matemático del FSE es más complicado que para un ecualizador síncrono y, por tanto, no se trata aquí.[10]

4.10 *Ecualización adaptable*

En esta sección formulamos un algoritmo simple pero efectivo para la ecualización adaptable de un canal lineal de características desconocidas. La figura 4.28 muestra la estructura de un ecualizador síncrono adaptable, el cual incorpora la acción del filtrado de acoplamiento. El algoritmo que se utiliza para ajustar los coeficientes del ecualizador supone la disponibilidad de una respuesta deseada. Una de las primeras reacciones a la disponibilidad de una réplica de la señal transmitida es: si dicha señal se dispone en el receptor, ¿por qué necesitamos ecualización adaptable? Para responder, advertimos primero que un canal telefónico característico cambia poco durante una llamada de datos promedio. Por tanto, antes de la transmisión de datos, el ecualizador se ajusta bajo la guía de una *secuencia de entrenamiento* transmitida a través del canal. Una versión sincronizada de esta secuencia

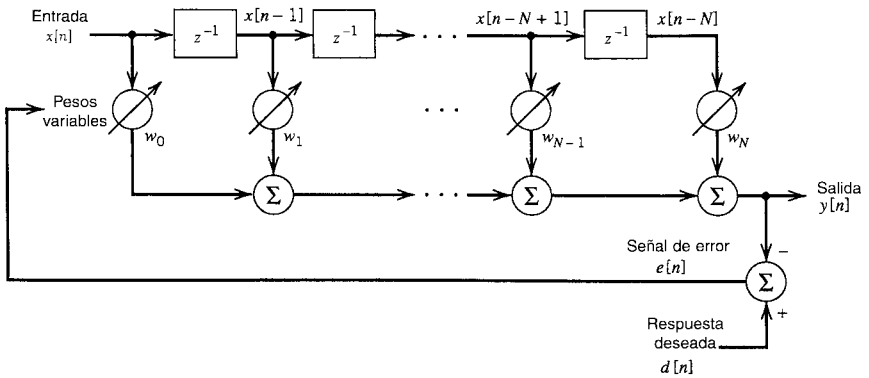


FIGURA 4.28 Diagrama de bloques de un ecualizador adaptable.

de adiestramiento se genera en el receptor, donde (después de un corrimiento de tiempo igual al retardo de la transmisión por el canal) se aplica al ecualizador como la respuesta deseada. Suele usarse en la práctica una secuencia de entrenamiento denominada de *pseudo-ruido,* que consiste en una secuencia periódica determinística con características similares al ruido. Se utilizan dos generadores idénticos de secuencia de pseudo-ruido, uno en el transmisor y el otro en el receptor. Cuando se completa el proceso de entrenamiento, el generador de secuencias de pseudo-ruido se desconecta y el ecualizador adaptable está listo para la transmisión de datos normal. La descripción detallada de los generadores de secuencia de pseudo-ruido se presenta en el capítulo 7.

■ ALGORITMO DE MÍNIMOS CUADRADOS (REENCUENTRO)

Para simplificar las cuestiones de notación, dejamos

$$x[n] = x(nT)$$
$$y[n] = y(nT)$$

Entonces, la salida $y[n]$ del ecualizador de retardo de línea en derivación en respuesta a la secuencia de entrada $\{x[n]\}$ se define mediante la suma de convolución discreta (vea la figura 4.28)

$$y[n] = \sum_{k=0}^{N} w_k x[n-k] \tag{4.112}$$

donde w_k es el peso de la k-ésima derivación y $N + 1$ es el número total de derivaciones. Los pesos de derivación constituyen los coeficientes de filtro adaptables. Suponemos que la secuencia de entrada $\{x[n]\}$ tiene energía finita. Hemos utilizado una notación para los pesos del ecualizador en la figura 4.28 que es diferente de la notación correspondiente en la figura 4.27 para subrayar el hecho de que el ecualizador de la figura 4.28 también incorpora filtrado de acoplamiento.

Es posible que la adaptación se consiga observando el error entre la forma del pulso deseada y la forma del pulso real a la salida del filtro, medida en los instantes de muestreo y utilizando después este error para estimar la duración en la cual los pesos de derivación del filtro deben variarse a fin de aproximarse a un conjunto de valores óptimo. Para la adaptación podemos utilizar un criterio que se basa en la minimización de la *distorsión pico* definida como la interferencia de intersímbolo en el peor caso a la salida del ecualizador. El desarrollo de un ecualizador adaptable al utilizar un criterio de este tipo se basa en el concepto de forzamiento cero, que se describe brevemente en la sección 4.9. Sin embargo, el ecualizador es óptimo sólo cuando la distorsión pico a su entrada es menor que 100% (es decir, la interferencia de intersímbolo no es demasiado severa). Un mejor método consiste en utilizar un criterio de error cuadrático medio, que es más general en la aplicación; asimismo, un ecualizador adaptable basado en un criterio de error cuadrático medio parece ser menos sensible a las perturbaciones de temporización que uno fundamentado en el criterio de distorsión pico. En consecuencia, a continuación utilizamos el criterio del error cuadrático medio para deducir el algoritmo de ecualización adaptable.

Sea $a[n]$ la *respuesta deseada* definida como la representación polar del n-ésimo símbolo binario que se transmite. Sea $e[n]$ la *señal de error* definida como la diferencia entre la respuesta deseada $a[n]$ y la respuesta real $y[n]$ del ecualizador, según indica

$$e[n] = a[n] - y[n] \tag{4.113}$$

En el *algoritmo de mínimos cuadrados*[11] para la ecualización adaptable, la señal de error $e[n]$ activa los ajustes aplicados a los pesos de derivación individuales del ecualizador cuando el algoritmo procede de una interacción a la siguiente. En la sección 3.13 se presentó una deducción del algoritmo de mínimos cuadrados para la predicción adaptable. Al reorganizar la ecuación (3.72) en una forma más general, es posible plantear la fórmula para el algoritmo de mínimos cuadrados en las siguientes palabras:

$$\begin{pmatrix}\text{Valor actualizado}\\ \text{del } k\text{-ésimo peso}\\ \text{en derivación}\end{pmatrix}=\begin{pmatrix}\text{Valor anterior}\\ \text{del } k\text{-ésimo peso}\\ \text{en derivación}\end{pmatrix}+\begin{pmatrix}\text{Parámetro}\\ \text{del tamaño}\\ \text{del escalón}\end{pmatrix}\bullet\begin{pmatrix}\text{Señal de entrada}\\ \text{aplicada al}\\ k\text{-ésimo peso}\\ \text{en derivación}\end{pmatrix}\begin{pmatrix}\text{Señal}\\ \text{de error}\end{pmatrix} \tag{4.114}$$

Sea μ el parámetro de tamaño de escalón. De la figura 4.28 vemos que la señal de entrada aplicada al k-ésimo peso de derivación en el tamaño del escalón n es $x[n-k]$. Por tanto, utilizando $\hat{w}_k(n)$ como el valor anterior del k-ésimo peso de derivación al tiempo del escalón n, el valor actualizado de este peso en derivación al tiempo del escalón $n+1$ está, de acuerdo con la ecuación (4.114), definido por

$$\hat{w}_k[n+1]=\hat{w}_k[n]+\mu x[n-k]e[n], \quad k=0,1,\ldots,N \tag{4.115}$$

donde

$$e[n]=a[n]-\sum_{k=0}^{N}\hat{w}_k[n]x[n-k] \tag{4.116}$$

Estas dos ecuaciones constituyen el *algoritmo de mínimos cuadrados para la ecualización adaptable.* Advierta que la longitud del ecualizador adaptable en la figura 4.28 no se confunde con la longitud del ecualizador de la figura 4.27.

Podemos simplificar la formulación del algoritmo de mínimos cuadrados utilizando la notación matricial. Dejemos que el vector $\mathbf{x}[n]$ de $(N+1)$ por 1 denote las entradas de derivación del ecualizador:

$$\mathbf{x}[n]=[x[n],\ldots,x[n-N+1],x[n-N]]^T \tag{4.117}$$

donde el exponente T indica la transposición de matriz. De modo correspondiente, dejemos que el vector $\hat{\mathbf{w}}[n]$ de $(N+1)$ por 1 denote los pesos de derivación del ecualizador:

$$\hat{\mathbf{w}}[n]=[\hat{w}_0[n],\hat{w}_1[n],\ldots,\hat{w}_N[n]]^T \tag{4.118}$$

Es posible entonces utilizar la notación de matrices para reformar la suma de convolución de la ecuación (4.112) en la forma compacta

$$y[n]=\mathbf{x}^T[n]\hat{\mathbf{w}}[n] \tag{4.119}$$

donde $\mathbf{x}^T[n]\hat{\mathbf{w}}[n]$ se conoce como el producto interior de los vectores $\mathbf{x}[n]$ y $\hat{\mathbf{w}}[n]$. Es posible resumir el algoritmo de mínimos cuadrados para la ecualización adaptable de la siguiente manera:

1. Se asignan valores iniciales al algoritmo dejando $\hat{\mathbf{w}}[1]=0$ (es decir, se igualan a cero todos los pesos de derivación del ecualizador en $n=1$, lo cual corresponde al tiempo $t=T$).
2. Para $n=1, 2, \ldots,$ se calcula

$$\begin{aligned}y[n]&=\mathbf{x}^T[n]\hat{\mathbf{w}}[n]\\ e[n]&=a[n]-y[n]\\ \hat{\mathbf{w}}[n+1]&=\hat{\mathbf{w}}[n]+\mu e[n]\mathbf{x}[n]\end{aligned}$$

   donde μ es el parámetro del tamaño del escalón.
3. Se continúa el cómputo interactivo hasta que el ecualizador alcanza un "estado estable", por el cual entendemos que el error cuadrático medio real del ecualizador alcanza esencialmente un valor constante.

El algoritmo de mínimos cuadrados constituye un ejemplo de un sistema de realimentación, de acuerdo a como se ilustra en el diagrama de bloques de la figura 4.29, el cual corresponde al k-ésimo

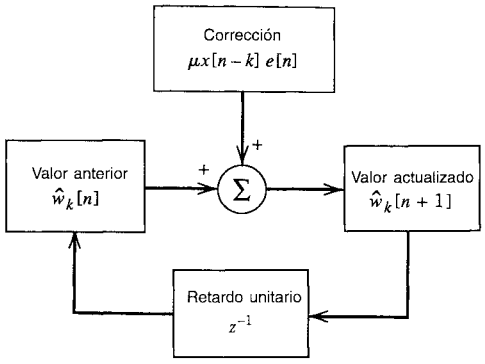


FIGURA 4.29 Representación gráfica del flujo de señales del algoritmo de mínimos cuadrados que incluye el k-ésimo peso de derivación.

coeficiente de filtro. Por tanto, es posible que el algoritmo diverja (es decir, para que el ecualizador adaptable se vuelva inestable). Desafortunadamente, el comportamiento convergente del algoritmo de mínimos cuadrados resulta difícil de analizar. A pesar de todo, siempre que se asigne un valor pequeño al parámetro μ del tamaño del escalón, encontramos que después de un gran número de interacciones el comportamiento del algoritmo de mínimos cuadrados es casi similar al del *algoritmo de descenso pronunciado*, que utiliza el gradiente real en vez de la estimación de ruido para el cálculo de los pesos de derivación. (El algoritmo de descenso pronunciado se explicó en la sección 3.13.)

■ OPERACIÓN DEL ECUALIZADOR

Existen dos modos de operación para un ecualizador adaptable, a saber, el de entrenamiento y el de decisión dirigida, como se muestra en la figura 4.30. Durante el *modo de entrenamiento*, como se explicó antes, se transmite una secuencia de pseudo-ruido conocida y se genera una versión sincronizada de ella en el receptor, donde (después de un corrimiento de tiempo igual a retardo de la transmisión) se aplica al ecualizador adaptable como la respuesta deseada; los pesos de derivación del ecualizador se ajustan de ese modo en concordancia con el algoritmo de mínimos cuadrados.

Cuando se completa el proceso de entrenamiento, el ecualizador adaptable se conmuta a su segundo modo de operación: el *de decisión dirigida*. En este modo de operación la señal de error está definida por

$$e[n] = \hat{a}[n] - y[n] \tag{4.120}$$

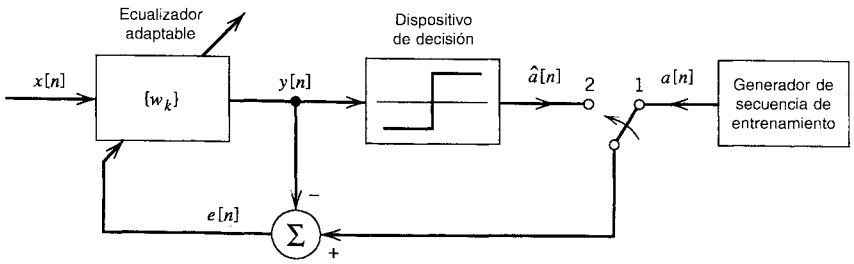


FIGURA 4.30 Ilustración de dos modos de operación de un ecualizador adaptable: en el modo de entrenamiento, el interruptor está en la posición 1; y en el modo de rastreo, se mueve a la posición 2.

donde $y[n]$ es la salida del ecualizador en el tiempo $t = nT$ y $\hat{a}[n]$ es la estimación correcta (no necesariamente final) del símbolo transmitido $a[n]$. Ahora bien, en operación normal las decisiones tomadas por el receptor son correctas con alta probabilidad. Esto significa que las estimaciones de error son correctas la mayor parte del tiempo, lo que permite que el ecualizador adaptable opere de modo satisfactorio. Asimismo, un ecualizador adaptable que opera en un modo de decisión dirigida es capaz de rastrear variaciones relativamente lentas de las características del canal.

Esto implica que cuanto mayor sea el parámetro del tamaño del escalón μ, tanto más rápida la capacidad de rastreo del ecualizador adaptable. Sin embargo, un parámetro del tamaño de escalón μ grande quizá resulte en un *error cuadrático medio de exceso* inaceptablemente alto, definido como aquella parte del valor cuadrático medio de la señal de error que excede del valor mínimo $J_{\min}$ alcanzable (mismo que se produce cuando los pesos de derivación están en sus ajustes óptimos). Por tanto, encontramos que la elección en la práctica de un valor adecuado del parámetro de tamaño de escalón μ implica establecer un compromiso entre el rastreo rápido y la reducción del error cuadrático medio de exceso.

Ecualización de decisión realimentada

Para obtener un mayor conocimiento de la ecualización adaptable considere un canal de banda base con respuesta al impulso denotada en su forma de muestreo por la secuencia $\{h[n]\}$, donde $h[n] = h(nT)$. La respuesta de este canal a una secuencia de entrada $\{x[n]\}$ a falta de ruido, está dada por la suma de convolución discreta

$$\begin{aligned} y[n] &= \sum_k h[k]x[n-k] \\ &= h[0]x[n] + \sum_{k<0} h[k]x[n-k] + \sum_{k>0} h[k]x[n-k] \end{aligned} \tag{4.121}$$

El primer término de la ecuación (4.121) representa el símbolo de datos deseado. El segundo término se debe a los *precursores* de la respuesta al impulso del canal que ocurren antes de la muestra principal $h[0]$ y se asocian con los símbolos de datos deseados. El tercer término corresponde a los *poscursores* de la respuesta al impulso del canal que ocurren después de la muestra principal $h[0]$. Los precursores y los poscursores de una respuesta al impulso del canal se ilustran en la figura 4.31. La idea de la *ecualización de decisión realimentada*[12] radica en utilizar las decisiones tomadas de los datos con base en los precursores de la respuesta al impulso del canal para ocuparse de los poscursores; para que la idea funcione, sin embargo, las decisiones tendrían evidentemente que ser correctas. Siempre y cuando se satisfaga esta condición, un ecualizador

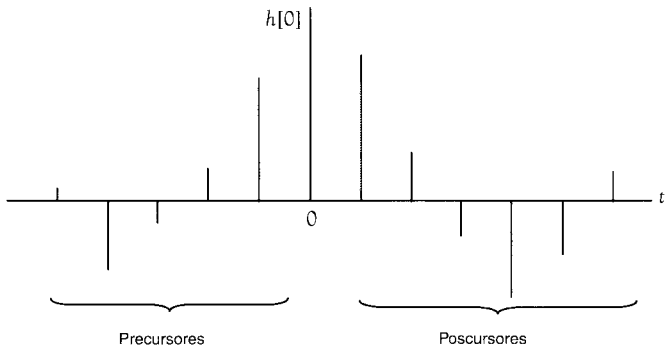


Figura 4.31 Respuesta al impulso de un canal en tiempo discreto, que describe los precursores y poscursores.

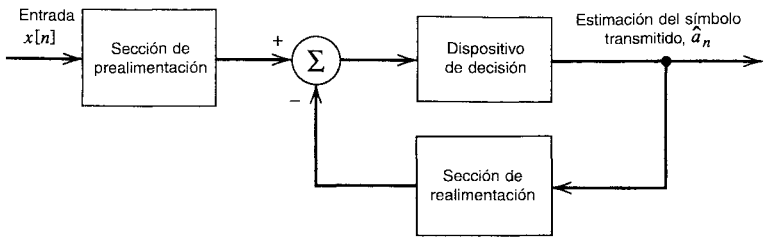


FIGURA 4.32 Diagrama de bloques de un ecualizador de decisión realimentado.

de decisión realimentado es capaz de proporcionar una mejora en el desempeño del ecualizador de retardo de línea en derivación.

Un *ecualizador de decisión realimentada* (DFE) consiste en una sección de prealimentación, una sección de realimentación y un dispositivo de decisión conectados en conjunto, como se muestra en la figura 4.32. La sección de prealimentación está integrada por un filtro de retardo de línea en derivación cuyas derivaciones están espaciadas de acuerdo con el recíproco de la velocidad de transmisión de señales. La secuencia de datos que se va a ecualizar se indica en esta sección. La sección de realimentación está compuesta por otro filtro de retardo de línea en derivación cuyas derivaciones se espacían también de acuerdo con el recíproco de la velocidad de transmisión de señales. La entrada que se aplica a la sección de realimentación corresponde a la decisión realizada sobre los símbolos de la secuencia de entrada que se detectaron previamente. La función de la sección de realimentación es sustraer de las estimaciones de muestras futuras la parte de la interferencia de intersímbolo producida por los símbolos detectados con anterioridad.

Advierta que la inclusión del dispositivo de decisión en el lazo de realimentación hace que el ecualizador sea intrínsecamente *no lineal* y, en consecuencia, más difícil de analizar que un ecualizador de retardo de línea en derivación ordinario. A pesar de eso, el criterio del error cuadrático medio puede utilizarse para obtener una optimización tratable matemáticamente de un ecualizador de decisión realimentado. En realidad, es posible utilizar el algoritmo de mínimos cuadrados para adaptar en forma conjunta tanto los pesos de derivación prealimentados como los pesos de derivación de realimentación sobre una señal de error *común*; vea el problema 4.37.

Con base en las evaluaciones comparativas extensivas de un ecualizador lineal y un ecualizador de decisión realimentada de los que se informa en la literatura,[13] es posible indicar que cuando la respuesta en frecuencia de un canal lineal está caracterizada por varias distorsiones de amplitud o cortes de amplitud relativamente agudos, el ecualizador de decisión realimentado ofrece una mejora importante en el desempeño con respecto a un ecualizador lineal para un número igual de derivaciones. Se presupone aquí que las decisiones de realimentación en el DFE son en su totalidad correctas. Para un ejemplo de un corte agudo de amplitud, vea la respuesta en frecuencia del canal telefónico que se describe en la figura 8 del capítulo de antecedentes y preliminares.

A diferencia de un ecualizador lineal, uno de decisión realimentada sufre de la *propagación de errores*. Sin embargo, a pesar del hecho de que DFE es un sistema realimentado, la propagación de errores no persistirá indefinidamente. Más bien, los errores de decisión tienden a ocurrir en *ráfaga*. Para justificar este tipo de comportamiento, ofrecemos el siguiente razonamiento intuitivo:[14]

- Sea L el número de derivaciones en la sección de realimentación de un DFE. Después de la secuencia de L decisiones consecutivas correctas, se desecharán todos los errores de decisión en la sección de realimentación. Esto indica un propagación de errores de duración finita.
- Cuando se comete un error de decisión, la probabilidad de que la siguiente decisión empiece a ser demasiado errónea claramente no es peor que 1/2.
- Sea K la duración de la propagación de errores; es decir, el número de símbolos necesarios para hacer L decisiones correctas consecutivas. Entonces la velocidad de error promedio es $(K/2)P_0$, donde $K/2$ es el número promedio de errores producido por un solo error de decisión, y P_0 es la probabilidad de error, dado que las L decisiones pasadas fueron todas correctas.

- En un experimento de lanzamiento de una moneda sin alterar, el número de lanzamientos, K, necesita que L caras sucesivas (al representar la inexistencia de errores) se vuelvan $2(2^L - 1)$.

Por tanto, se concluye que el efecto de la propagación de errores en un ecualizador de decisión realimentado radica en aumentar la velocidad de error promedio en un factor aproximadamente igual a 2^L, en comparación con la probabilidad de realizar el primer error. Por ejemplo, para $L = 3$ la velocidad de error promedio se incrementa en menos de un orden de magnitud debido a la propagación de errores.

4.11 *Experimentos de computadora: patrones del ojo*

En las secciones anteriores de este capítulo hemos explicado varias técnicas para tratar los efectos del ruido del canal y de la interferencia de intersímbolo en el desempeño de un sistema de transmisión de pulsos de banda base. En el análisis final, lo que en realidad importa es cómo evaluar el efecto combinado de estos deterioros sobre el desempeño total del sistema en un ambiente operativo. Una herramienta experimental para una evaluación exhaustiva de este tipo es el llamado *patrón del ojo*, el cual se define como la superposición sincronizada de todas las realizaciones posibles de la señal de interés (por ejemplo, señal recibida, salida del receptor) considerada dentro de un intervalo de transmisión de señales particular. El patrón del ojo deriva su nombre del hecho de que se asemeja al ojo humano en el caso de ondas binarias. La región interior del patrón del ojo recibe el nombre de *apertura ocular*.

Un patrón del ojo proporciona una gran cantidad de información útil acerca del desempeño de un sistema de transmisión de datos, como se describe en la figura 4.33. Específicamente, hacemos los siguientes enunciados:

- El ancho de la apertura ocular define el *intervalo del tiempo sobre el cual la señal recibida puede muestrearse sin error de la interferencia de intersímbolo;* es claro que el tiempo que se prefiere para el muestreo es el instante en el cual el ojo está abierto al máximo.
- La *sensibilidad del sistema a los errores de temporización* está determinada por la velocidad de cierre del ojo cuando se varía el tiempo de muestreo.
- La altura de la apertura ocular, a un tiempo de muestreo especificado, define el *margen de ruido* del sistema.

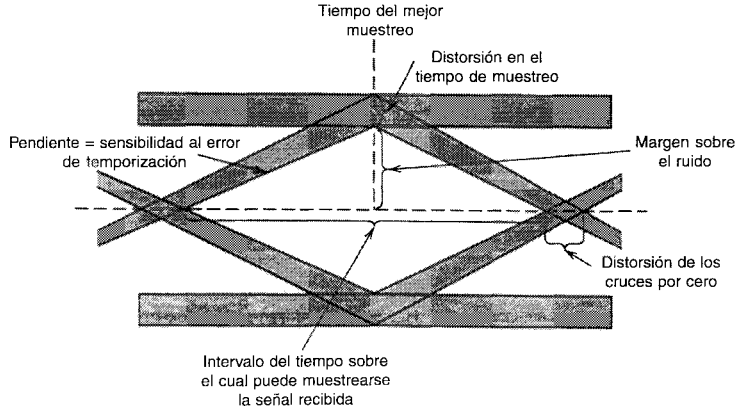


FIGURA 4.33 Interpretación del patrón del ojo.

Cuando es severo el efecto de la interferencia de intersímbolo, trazas de la porción superior del patrón del ojo cruzan por trazas de la porción inferior, con el resultado de que el ojo está por completo cerrado. En una situación tal, es posible evitar errores debido a la presencia combinada de la interferencia de intersímbolo y el ruido en el sistema.

En el caso de un sistema M-ario, el patrón del ojo contiene (M − 1) aperturas oculares apiladas verticalmente, donde M es el número de niveles de amplitud discretos utilizados para construir la señal transmitida. En un sistema lineal con datos verdaderamente aleatorios, todas estas aperturas del ojo serían idénticas.

En los siguientes dos experimentos utilizamos simulaciones de computadora para estudiar los patrones del ojo correspondientes a un sistema de transmisión PAM de banda base cuaternario (M = 4) en condiciones sin ruido, ruidosas y de banda limitada. El efecto de la no linealidad del canal sobre los patrones del ojo se discute en el problema 4.38.

Experimento 1: efecto del ruido del canal

La figura 4.34*a* muestra el diagrama del ojo del sistema bajo condiciones idealizadas: sin el ruido del canal y sin limitación del ancho de banda. Los símbolos de fuente que se utilizan se generan en forma aleatoria en una computadora, con la forma de pulso de coseno elevado. Los parámetros del sistema que se utilizan para la generación del diagrama del ojo son como sigue: ancho de banda de Nyquist $W = 0.5$ Hz, factor de caída $\alpha = 0.5$ y duración del símbolo $T = T_b \log_2 M = 2T_b$. Las aperturas en la figura 4.34 son perfectas, lo que indica la operación confiable del sistema. Advierta que esta figura tiene $M - 1 = 3$ aperturas.

Las figuras 4.34*b* y 4.34*c* muestran los diagramas del ojo para el sistema, pero esta vez con ruido de canal corrompiendo la señal recibida. Estas dos figuras se simularon para una relación señal a ruido SNR = 20 dB y 10 dB, respectivamente, con la SNR midiéndose a la salida del canal. Cuando SNR = 20 dB, el efecto del ruido del canal difícilmente se distingue en la figura 4.34*b*, pero cuando su valor es de SNR = 10 dB, las aperturas del diagrama del ojo en la figura 4.34*c* son apenas visibles.

Experimento 2: efecto de la limitación del ancho de banda

Las figuras 4.35*a* y 4.35*b* muestran los diagramas oculares para el sistema cuaternario utilizando los mismos parámetros que antes, aunque en este caso bajo la condición de ancho de banda limitado y un canal sin ruido. De manera específica, el canal se modela ahora por medio de un *filtro* de *Butterworth* pasobajas, cuya respuesta en magnitud al cuadrado está definida por

$$|H(f)|^2 = \frac{1}{1 + (f/f_0)^{2N}}$$

donde N es el *orden* del filtro y f_0 es la frecuencia de corte de 3 dB. Para el experimento de computadora que se describe en la figura 4.35*a* se utilizan los siguientes valores:

$$N = 25 \text{ y } f_0 = 0.975 \text{ Hz}$$

El ancho de banda requerido para el sistema de transmisión PAM se calcula como

$$B_T = W(1 + \alpha) = 0.75 \text{ Hz}$$

Aunque el ancho de banda del canal (es decir, la frecuencia de corte) es mayor que la absolutamente necesaria, su efecto en la banda de paso se observa como una disminución en el tamaño de las aperturas del ojo en comparación con las de la figura 4.34*a*. En vez de valores nítidos en el tiempo $t = 1$ s (como se muestra en la figura 4.34*a*), ahora hay una región borrosa.

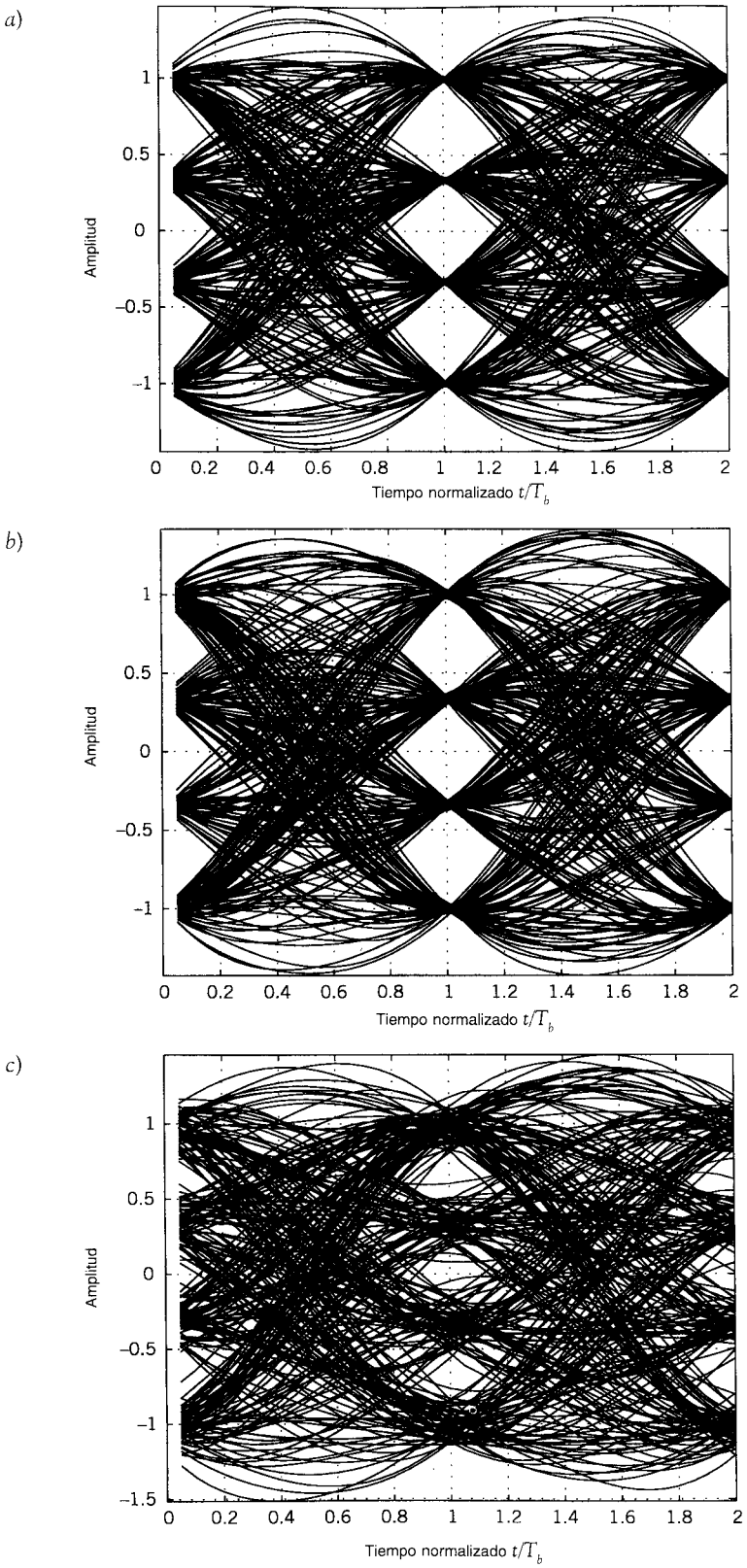


FIGURA 4.34 *a*) Diagrama del ojo para un sistema cuaternario sin ruido. *b*) Diagrama del ojo para un sistema cuaternario con NSR = 20 dB. *c*) Diagrama del ojo para un sistema cuaternario con NSR = 10 dB.

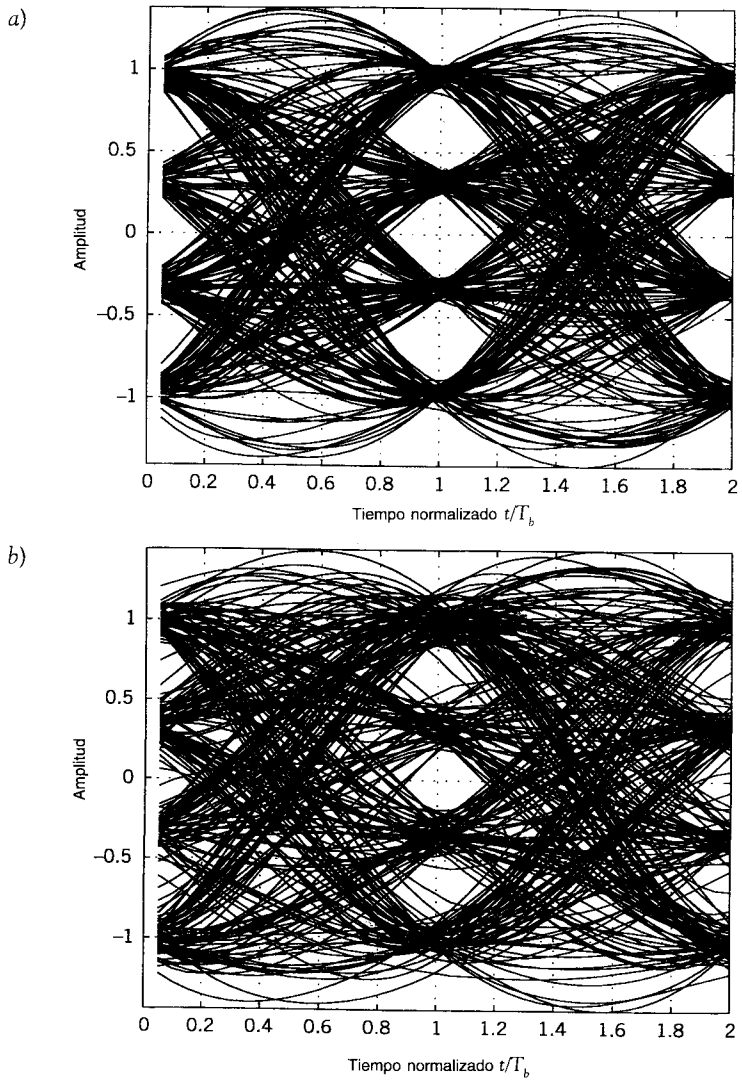


FIGURA 4.35 *a*) Diagrama del ojo para un sistema cuaternario de banda limitada sin ruido: frecuencia de corte $f_0 = 0.975$ Hz. *b*) Diagrama del ojo para un sistema cuaternario de banda limitada sin ruido: frecuencia de corte $f_0 = 0.5$ Hz.

En la figura 4.35*b* el ancho de banda del canal se reduce aún más al modelar el canal como un filtro de Butterworth pasobajas con $N = 25$ y $f_0 = 0.5$ Hz. El efecto del ancho de banda reducido del canal es reducir aún más el margen al cual se abren los ojos.

4.12 *Resumen y discusión*

En este capítulo estudiamos los efectos del ruido del canal y de la interferencia de intersímbolo sobre el desempeño de los sistemas de transmisión de pulsos de banda base. La interferencia de intersímbolo (ISI) es diferente del ruido en cuanto a que ésta es una forma de interferencia *dependiente de la señal* que se produce debido a las desviaciones en la respuesta en frecuencia de un canal con respecto a la del filtro pasobajas ideal (canal de Nyquist); dicha interferencia desaparece cuando la señal transmitida se corta. El resultado de estas desviaciones es que el pulso recibido correspondiente a un símbolo

de datos particular resulta afectado por los extremos de las colas de los pulsos que representan a los símbolos previos y los extremos delanteros de los pulsos que representan a los símbolos subsecuentes.

Dependiendo de la relación señal a ruido recibida, es posible distinguir tres situaciones diferentes que pueden presentarse en los sistemas de transmisión de pulsos de banda base para canales con características fijas:

1. *El efecto de la interferencia de intersímbolo es despreciable en comparación con la del ruido del canal.* El procedimiento apropiado en este caso consiste en utilizar un filtro de acoplamiento, el cual es el filtro lineal invariante con el tiempo óptimo para maximizar la relación señal a ruido del pulso pico.

2. *La relación señal a ruido recibida es suficientemente alta para ignorar el efecto del ruido del canal.* En este caso necesitamos asegurarnos contra los efectos de la interferencia de intersímbolo sobre la reconstrucción de los datos transmitidos en el receptor. En particular, debe ejercerse el control sobre la forma del pulso que se recibe. Este objetivo de diseño puede lograrse de una de dos maneras diferentes:

   - Utilizando un espectro de coseno elevado para la respuesta en frecuencia completa del sistema de transmisión de pulsos de banda base.
   - Utilizando codificación de nivel correlativo o transmisión de señales de respuesta parcial que añada de una manera controlada la interferencia de intersímbolo a la señal transmitida.

3. *Son importantes tanto la interferencia de intersímbolo como el ruido.* En relación con una solución tratable matemáticamente en esta más difícil situación, es posible recurrir al criterio del error cuadrático medio. El receptor lineal óptimo resultante recibe el nombre de receptor de error cuadrático medio mínimo (mmse). Éste consiste en una conexión en cascada de un filtro de acoplamiento y un ecualizador transversal lineal (retardo de línea en derivación).

Sin embargo, cuando el canal es aleatorio en el sentido de ser un ensamble de posibles realizaciones físicas, lo que es con frecuencia el caso en un ambiente de telecomunicaciones, tal vez no sea adecuado el uso de diseños de filtro fijo basados en las características promedio del canal. En situaciones de este tipo, el método que se prefiere implica el uso de un ecualizador adaptable, cuyo propósito es compensar en forma automática las variaciones en la respuesta en frecuencia del canal durante el curso de la transmisión de datos. El uso combinado de un filtro de retardo de línea en derivación y del algoritmo de mínimos cuadrados para ajustar los pesos en derivación, proporciona la base de un método simple pero altamente efectivo para poner en práctica el ecualizador adaptable. Un dispositivo de estas características es capaz de enfrentar los efectos combinados de la interferencia de intersímbolo y del ruido del receptor en un ambiente no estacionario. Su valor práctico se encuentra en el hecho de que casi cualquier módem (modulador-demodulador) en actual uso comercial para la transmisión de datos digitales por un canal telefónico de voz utiliza un ecualizador adaptable como parte integral.

Otra aplicación importante del filtrado adaptable se encuentra en el diseño de los canceladores de eco que conforman un componente crítico de los transceptores para líneas digitales de suscriptor. Por lo común, una línea digital de suscriptor utiliza un par trenzado como el medio de transmisión, el mismo que se utiliza en los canales telefónicos ordinarios. Sin embargo, a diferencia de los canales telefónicos, las líneas digitales de suscriptor se diseñan para ofrecer una capacidad de transmisión de datos digitales de alta velocidad entre una red digital y las plantas de suscriptores, con una velocidad de datos de 64 kb/s y valores superiores.

NOTAS Y BIBLIOGRAFÍA

1. Los libros clásicos sobre la transmisión de pulsos por la banda base son Lucky, Salz y Weldon [97] y Sunde [151]. Para un tratamiento detallado de los diferentes aspectos del tema, consulte Gitlin, Hayes y Weinstein [52], Proakis [123] y Benedetto, Biglieri y Castellani [12].

2. La caracterización de un filtro de acoplamiento la dedujo por primera vez North en un informe secreto (RCA Laboratories Report PTR-6C en junio de 1943), el cual se publicó veinte años después: vea el artículo de North [330]. Un resultado similar lo obtuvieron de forma independiente Van Vleck y Middleton, quienes acuñaron el término *filtro de acoplamiento*: vea el artículo de Van Vleck y Middleton [388]. Con respecto a una revisión del material acerca del filtro de acoplamiento y sus propiedades, vea los artículos de Turin [382, 383].

3. La *función de error* denotada por $\text{fer}(u)$, se define de varias maneras diferentes en la literatura. Utilizaremos la siguiente definición:

$$\text{fer}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-z^2)\, dz$$

La función de error tiene dos propiedades útiles:

i) $\text{fer}(-u) = -\text{fer}(u)$
Esto se conoce como la *relación de simetría.*

ii) Cuando u tiende a infinito, $\text{fer}(u)$ tiende a la unidad; esto es

$$\frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-z^2)\, dz = 1$$

La *función de error complementaria* se define por medio de

$$\text{ferc}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-z^2)\, dz$$

que se relaciona con la función de error del modo siguiente:

$$\text{ferc}(u) = 1 - \text{fer}(u)$$

La tabla A6.6 proporciona valores de la función de error $\text{fer}(u)$ para u en el intervalo de 0 a 3.3.

Para valores grandes y positivos de u, tenemos dos cotas simples para $\text{ferc}(u)$, una inferior y otra superior, como se indica mediante

$$\frac{\exp(-u^2)}{\sqrt{\pi u}}\left(1 - \frac{1}{2u^2}\right) < \text{ferc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi u}}$$

La función de error complementaria proporciona la base para una formulación compacta de la probabilidad del error de símbolo, como se explica en la sección 4.3. Otra sección que también suele usarse en la literatura para este fin es la función Q. Considere una variable aleatoria gaussiana *estandarizada* X de media cero y varianza unitaria. La probabilidad de que un valor observado de la variable aleatoria X será mayor que v está dada por la *función de Q:*

$$Q(v) = \frac{1}{\sqrt{2\pi}} \int_v^\infty \exp\left(-\frac{x^2}{2}\right) dx$$

La función de Q define el área bajo la cola gaussiana estandarizada. La función de Q se relaciona con la función de error complementaria como

$$Q(v) = \frac{1}{2}\text{ferc}\left(\frac{v}{\sqrt{2}}\right)$$

Inversamente, dejando $u = v/\sqrt{2}$, tenemos

$$\text{ferc}(u) = 2Q(\sqrt{2}u)$$

4. El criterio descrito en la ecuación (4.49) o en la ecuación (4.53) lo formuló Nyquist por primera vez en el estudio de la teoría de la transmisión telegráfica; el artículo que publicó dicho autor en 1928 constituye un clásico. En la literatura, este criterio se conoce como *primer criterio de Nyquist*. En su artículo de 1928, Nyquist describió otro método, conocido en la literatura como *segundo criterio de*

Nyquist. El segundo método utiliza los instantes de transición entre símbolos distintos en la señal recibida más que muestras centradas. Una discusión sobre el primero y segundo criterios se presenta en Bennett ([13], pp. 78-92) y en el artículo de Gibby y Smith [261]. Un tercer criterio que se atribuye a Nyquist se explica en Sunde [151]; consulte también los artículos de Pasupathy [337] y Sayar y Pasupathy [364].

5. La codificación de nivel correlativo y la transmisión de señales de respuesta parcial son sinónimos; ambos términos se usan en la literatura. La idea de codificación correlativa la originó Lender [302]. El trabajo de este autor lo generalizó Kretzmer [298] para la transmisión de datos binarios. Para mayores detalles acerca de las técnicas de codificación correlativa, vea el libro de Gitlin, Hayes y Weinstein [52]; revise también los artículos de Pasupathy [338], Kabal y Pasupathy [291] y Sousa y Pasupathy [377].

6. El material de las líneas digitales de suscriptor que se presenta en la sección 4.8 se basó en los artículos de Lin y Tzeng [305] y Lechleider [300], así como los libros de Starr, Cioffi y Silverman [147] y Chen [29].

7. Con respecto a una discusión de los códigos de línea para lazos digitales de suscriptor, vea Gitlin y otros [52].

8. Ericson demuestra que para cada criterio de desempeño "razonable", el receptor óptimo puede realizarse como un filtro de acoplamiento seguido por un ecualizador de retardo de línea en derivación, como se muestra en la figura 4.27. Aparte del criterio del error cuadrático medio que se consideró en la sección 4.9, entre los criterios de desempeño razonable de interés se incluyen:

   i) Minimización de la probabilidad del error de símbolo.
   ii) Ecualización de forzamiento cero (para reducir la interferencia de intersímbolo a cero), seguido por minimización de la probabilidad del error de símbolo considerando esta restricción.
   iii) Minimización de la relación señal a ruido en los instantes de muestreo.

   El criterio *i* es el método más natural para la optimización de un receptor lineal; este método, seguido en Aaron y Tufts [179], resulta, desafortunadamente, complicado. El criterio *ii*, resultado de Lucky y otros [97], es un método mucho más simple. El criterio *iii* se debe a George [259].

9. La ecualización adaptable de canales telefónicos fue propuesta por Lucky [312, 313]. Desde entonces se han publicado en la literatura diversos esquemas de ecualización adaptable, los cuales proporcionan ecualización para sistemas síncronos de transmisión de datos específicos. Con respecto a artículos restrospectivos sobre la ecualización adaptable, consulte Proakis [345] y Qureshi [346, 347]. La ecualización adaptable se explica también en detalle en los libros de Gitlin, Hayes y Weinstein ([52], capítulo 8) y Proakis ([345], capítulo 6).

10. Parece que el primer trabajo sobre los ecualizadores espaciados fraccionalmente fue iniciado por Brady. Otras contribuciones al tema incluyeron el trabajo subsecuente de Ungerboeck, y Gitlin y Weinstein. Una discusión detallada de los ecualizadores espaciados fraccionalmente se presenta en Gitlin y otros [52].

11. El algoritmo de mínimos cuadrados fue originado por Widrow y Hoff, Jr. Para un análisis de convergencia detallado del algoritmo de mínimos cuadrados, consulte Haykin ([62], capítulo 9), y Widrow y Stearns ([167], capítulo 6).

12. La ecualización de decisión realimentada fue descrita primero en un informe de Austin [192]. La optimización del ecualizador de decisión realimentado para un error cuadrático medio mínimo la llevó a cabo por primera vez Monsen [325]. Una explicación accesible de la ecualización de decisión realimentada se presenta en el libro de Gitlin, Hayes y Weinstein ([52], pp. 500-510).

    Tomlinson [380] y Harashima y Miyakawa [277] describen un dispositivo para eliminar la propagación de errores en un ecualizador de decisión realimentado. El dispositivo, conocido como el *precodificador de Tomlinson-Harashima*, aparece en el transmisor como un preprocesador para el modulador. La idea básica de este precodificador es mover la sección de realimentación en el ecualizador de decisión realimentado al transmisor donde es imposible realizar errores de decisión.

Sin embargo, esta modificación puede resultar en un aumento significativo de la potencia de transmisión; se utiliza un módulo aritmético para superar la mayor parte de este aumento de potencia.

13. Para la comparación del desempeño entre ecualizadores lineales y ecualizadores de decisión realimentados, vea Gitlin y otros [52] y Proakis [345].

14. La explicación intuitiva acerca de la propagación de errores en los ecualizadores de decisión realimentados que se presentó en la sección 4.10, se obtuvo de Gitlin y otros [52].

    Para una evaluación rigurosa de la probabilidad del error de símbolo P_e en un ecualizador de decisión realimentado con propagación de errores, vea Duttweiler y otros. En este artículo se demuestra que en la interferencia de intersímbolo para el peor caso, P_e, se multiplica por un factor de 2^L relativa a la probabilidad de error que se produce a falta de errores de decisión y en relaciones señal a ruido altas, donde L es el número de derivaciones en la sección de realimentación. El resultado que obtuvieron Duttweiler y otros proporciona la justificación teórica para los argumentos intuitivos que se presentan en la sección 4.10.

PROBLEMAS

Filtros de acoplamiento

4.1 Considere la señal $s(t)$ que se muestra en la figura P4.1.

a) Determine la respuesta al impulso de un filtro acoplado a esta señal y dibújela como función del tiempo.

b) Grafique la salida del filtro de acoplamiento como una función del tiempo.

c) ¿Cuál es el valor pico de la salida?

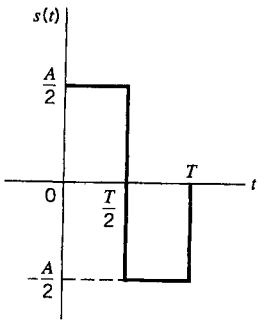


FIGURA P4.1

4.2 La figura P4.2*a* muestra un par de pulsos que son ortogonales entre sí sobre el intervalo $[0,T]$. En este problema investigamos el uso de este par de pulsos para estudiar un *filtro de acoplamiento bidimensional.*

a) Determine los filtros de acoplamiento para los pulsos $s_1(t)$ y $s_2(t)$ considerados en forma individual; para $s_1(t)$ el filtro es el mismo que el que se consideró en el problema 4.1.

b) Forme un filtro de acoplamiento bidimensional conectando en paralelo los dos filtros acoplados de la parte *a*, como se muestra en la figura 4.2*b*. Por consiguiente, demuestre que:

i) Cuando el pulso $s_1(t)$ se aplica a este filtro bidimensional, la respuesta del filtro de acoplamiento inferior es cero.

ii) Cuando se aplica el pulso $s_2(t)$ al filtro bidimensional, la respuesta del filtro de acoplamiento superior es cero.

Generalice los resultados de su investigación.

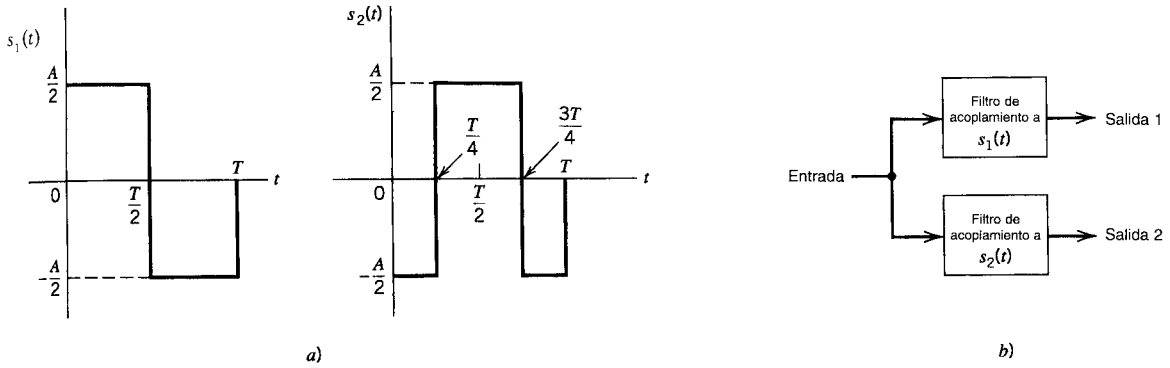


FIGURA P4.2

4.3 Considere un pulso rectangular definido por

$$g(t) = \begin{cases} A, & 0 \le t \le T \\ 0, & \text{en otro caso} \end{cases}$$

Se propone aproximar el filtro de acoplamiento para $g(t)$ mediante un filtro pasobajas ideal de ancho de banda B; la maximización de la relación señal a ruido del pulso pico es el objetivo principal.

a) Determine el valor óptimo de B para el cual el filtro pasobajas ideal proporciona la mejor aproximación al filtro de acoplamiento.

b) ¿En cuántos decibeles el filtro pasobajas ideal empeora con respecto al filtro de acoplamiento?

4.4 En este problema exploramos otro método para la realización aproximada de un filtro de acoplamiento, esta vez utilizando el simple filtro pasobajas resistencia-capacitancia (RC) que se muestra en la figura P4.4. La respuesta en frecuencia de este filtro es

$$H(f) = \frac{1}{1 + jf/f_0}$$

donde $f_0 = 1/2\pi RC$. La señal de entrada $g(t)$ es un pulso rectangular de amplitud A y duración T. Se requiere optimizar la selección de la frecuencia de corte de 3 dB, f_0, del filtro, de manera que la relación señal a ruido del pulso pico a la salida del filtro se haga máxima. Con este objetivo en mente, demuestre que el valor óptimo de f_0 es $0.2/T$, para el cual la pérdida de relación señal a ruido comparada con el filtro de acoplamiento es aproximadamente de 1 dB.

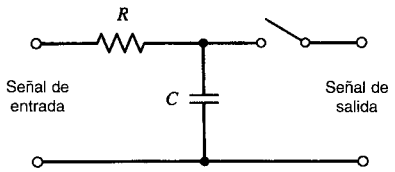


FIGURA P4.4

Probabilidad del cálculo del error

4.5 La fórmula para el umbral óptimo en el receptor de la figura 4.4 está dada, en general, por la ecuación (4.37). Analice, en términos gráficos, cómo afecta esta elección a las contribuciones de los dos

términos en la ecuación (4.35) a la probabilidad promedio del error de símbolo, P_e, considerando los siguientes dos casos:

a) $p_0 > p_1$

b) $p_1 < p_0$

donde p_0 y p_1 son las probabilidades *a priori* de los símbolos 0 y 1, respectivamente.

4.6 En un sistema PCM binario, los símbolos 0 y 1 tienen *probabilidades a priori* p_0 y p_1, respectivamente. La función de densidad de probabilidad condicional de la variable aleatoria Y (con valor de muestra y) obtenida mediante el muestreo de la salida del filtro de acoplamiento en el receptor de la figura 4.4 al final del intervalo de transmisión de la señal, dado que se transmitió el símbolo 0, se denota por medio de $f_Y(y|0)$. De manera similar, $f_Y(y|1)$ denota la función densidad de probabilidad condicional de Y, dado que se transmitió el símbolo 1. Considere que λ denota el umbral utilizado en el receptor, de modo que si el valor de muestreo y supera a λ, el receptor decide a favor del símbolo 1; de otro modo, decide a favor del símbolo 0. Demuestre que el umbral óptimo λ_{opt}, por el cual la probabilidad promedio de error es un mínimo, está dada por la solución de

$$\frac{f_Y(\lambda_{opt}|1)}{f_Y(\lambda_{opt}|0)} = \frac{p_0}{p_1}$$

4.7 Un sistema PCM binario que utiliza transmisión de señales polares sin retorno a cero opera justo arriba del umbral de error con una probabilidad de error promedio igual a 10^{-6}. Suponga que se duplica la velocidad de transmisión de señales. Determine un nuevo valor de la probabilidad de promedio de error. Puede utilizar la tabla A6.6 para evaluar la función de error complementaria.

4.8 Una señal en tiempo continuo se muestrea y luego se transmite como una señal PCM. La variable aleatoria a la entrada del dispositivo de decisión en el receptor tiene una varianza de 0.01 volts^2.

a) Suponiendo el uso de transmisión de señales polares sin retorno a cero, determine la amplitud del pulso que debe transmitirse para que la tasa de error promedio no supere 1 bit en 10^8 bits.

b) Si la presencia añadida de interferencia ocasiona que la tasa de error aumente a 1 bit en 10^6, ¿cuál es la varianza de la interferencia?

4.9 Una onda PCM binaria utiliza transmisión de señales unipolares sin retorno a cero para transmitir los símbolos 1 y 0; el símbolo 1 está representado por un pulso rectangular de amplitud A y duración T_b. El ruido del canal se modela como aditivo, blanco y gaussiano, con media cero y densidad espectral de potencia $N_0/2$. Suponiendo que el símbolo 1 y el 0 ocurren con igual probabilidad, determine una expresión para la probabilidad promedio del error a la salida del receptor, utilizando un filtro de acoplamiento como se describe en la sección 4.3.

4.10 Repita el problema 4.9 para el caso de transmisión de señales unipolares con retorno a cero, donde el símbolo 1 se representa mediante un pulso de amplitud A y duración $T_b/2$ y el símbolo 0 se representa por la no transmisión de pulso.

De acuerdo con ello, demuestre que este tipo de unipolaridad de la transmisión de señales requiere el doble de la potencia promedio de la transmisión unipolar sin retorno a cero (es decir, activación–desactivación) para la misma probabilidad promedio del error de símbolo.

4.11 En este problema regresamos al receptor PCM de la figura 4.4, pero esta vez consideramos el uso de transmisión de señales bipolares sin retorno a cero, en cuyo caso la señal transmitida $s(t)$ se define por medio de

Símbolo binario 1: $s(t) = \pm A$ para $0 < t \leq T$

Símbolo binario 0: $s(t) = 0$, $0 < t \leq T$

Determine la probabilidad promedio del error de símbolo P_e para este receptor suponiendo que los números binarios 0 y 1 son de igual probabilidad.

Espectro del coseno elevado

4.12 El pulso sin retorno a cero de la figura P4.12 puede considerarse como una forma muy burda de un pulso de Nyquist. Compare las características espectrales de estos dos pulsos.

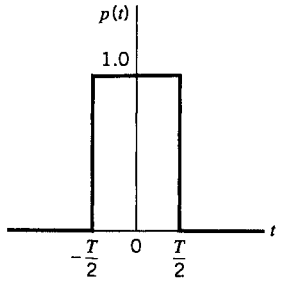


FIGURA P4.12

4.13 Determine la transformada inversa de Fourier de la función de frecuencia $P(f)$ definida en la ecuación (4.60).

4.14 Una señal analógica se muestrea, cuantiza y codifica en una onda PCM binaria. Las especificaciones del sistema PCM incluyen lo siguiente:

Velocidad de muestreo = 8 kHz

Número de niveles de representación = 64

La onda PCM se transmite sobre un canal de banda base utilizando modulación por amplitud de pulsos discreta. Determine el ancho de banda mínimo requerido para transmitir la onda PCM si se deja que cada pulso tome los siguientes números de niveles de amplitud: 2, 4 u 8.

4.15 Considere un sistema PAM binario de banda base que se diseña para tener un espectro de coseno elevado $P(f)$. El pulso que resulta $p(t)$ se define en la ecuación (4.62). ¿Cómo podría modificarse este pulso si se hubiera diseñado este sistema para tener una respuesta en fase lineal?

4.16 Una computadora produce datos binarios a razón de 56 kb/s. La salida de la computadora se transmite utilizando un sistema PAM binario de banda base que se diseña para tener un espectro de coseno elevado. Determine el ancho de banda de transmisión que se requiere para cada uno de los siguientes factores de caída:

$\alpha = 0.25, 0.5, 0.75, 1.0$.

4.17 Repita el problema 4.16, dado que cada conjunto de tres dígitos binarios sucesivos en la salida de la computadora se codifican en uno de ocho posibles niveles de amplitud, y que la señal resultante se transmite empleando un sistema PAM de ocho niveles diseñado para tener espectro de coseno elevado.

4.18 Una señal analógica se muestrea, cuantiza y codifica en una PCM binaria. El número de niveles de representación que se utiliza corresponde a 128. Se agrega un pulso de sincronización al final de cada palabra de código que representa a una muestra de la señal analógica. La onda PCM resultante se transmite por un canal de 12 kHz de ancho de banda utilizando un sistema PAM cuaternario con espectro de coseno elevado. El factor de caída es la unidad.

a) Determine la velocidad (b/s) a la cual se transmite la información por el canal.

b) Obtenga la velocidad a la cual se muestrea la señal analógica. ¿Cuál es el valor máximo posible para la componente de frecuencia más alta de la señal analógica?

4.19 Una onda PAM binaria se va a transmitir por un canal de banda base con un ancho de banda máximo absoluto de 75 kHz. La duración de los bits es de 10 μs. Determine el espectro de coseno elevado que satisface estos requerimientos.

Codificación de niveles correlativos

4.20 Las técnicas de transmisión de señales duobinaria, ternaria y bipolar tienen un rasgo común: todas emplean tres niveles de amplitud. ¿En qué difiere la técnica duobinaria de las otras dos?

4.21 La secuencia de datos binaria 001101001 se aplica a la entrada de un sistema duobinario.

a) Construya una salida de codificador duobinario y la correspondiente salida del receptor, sin un precodificador.

b) Suponga que debido a errores durante la transmisión, el nivel a la entrada del receptor que produce el segundo dígito se reduce a cero. Construya la nueva salida del receptor.

4.22 Repita el problema 4.21 suponiendo el empleo de un precodificador en el transmisor.

4.23 El esquema que se muestra en la figura P4.23 puede considerarse como un codificador diferencial (consistente en el sumador de módulo-dos y en el elemento de retardo de una unidad) conectado en cascada con una forma especial de codificador correlativo (compuesto por un elemento de retardo de una unidad y un sumador). En la figura P4.23 se presenta un único elemento de retardo, puesto que éste es común tanto para el codificador diferencial como para el codificador correlativo. En este codificador diferencial, una transición se representa mediante el símbolo 0 y la no transmisión por medio del símbolo 1.

a) Encuentre la respuesta en frecuencia y la respuesta al impulso de la parte del codificador correlativo del esquema que se muestra en la figura P4.23.

b) Muestre que este esquema puede utilizarse para convertir la representación activado-desactivado de una secuencia binaria (aplicada a la entrada) en una representación bipolar de la frecuencia a la salida. Usted puede ilustrar esta conversión considerando la frecuencia 010001101.

Para descripciones de la codificación activado-desactivado, bipolar y diferencial de secuencias binarias, vea la sección 3.7.

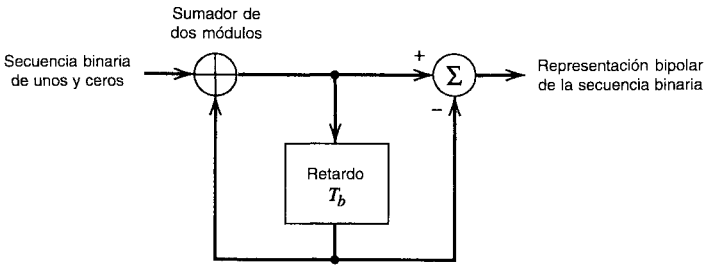


FIGURA P4.23

4.24 Considere una onda binaria aleatoria $x(t)$ en la cual los unos y ceros ocurren con igual probabilidad, los símbolos en periodos adyacentes son estadísticamente independientes, el símbolo 1 se representa mediante A volts y el símbolo 0 por medio de 0 volts. Esta onda binaria activado-desactivado se aplica al circuito de la figura P4.23.

a) Utilizando el resultado del problema 4.23, muestre que la densidad espectral de potencia de la onda bipolar $y(t)$ que aparece a la salida del circuito es igual a

$$S_Y(f) = T_b A^2 \operatorname{sen}^2(\pi f T_b)\ \operatorname{senc}^2(f T_b)$$

b) Grafique las densidades espectrales de potencia de las ondas binarias activado-desactivado y bipolar, y compárelas.

4.25 La secuencia de datos binaria 011100101 se aplica a la entrada de un sistema duobinario modificado.

a) Construya el codificador duobinario modificado y la correspondiente salida del receptor, sin un precodificador.

b) Supóngase que debido a un error durante la transmisión, el nivel producido por el tercer dígito se reduce a cero. Construya la nueva salida del receptor.

4.26 Repita el problema 4.25 suponiendo el uso de un precodificador en el transmisor.

Sistemas PAM M-arios

4.27 Considere un sistema M-ario utilizando M niveles de amplitud discretos. El modelo del receptor se muestra en la figura P4.27, cuya operación está determinada por las siguientes suposiciones:

a) La componente de la señal en la onda recibida es

$$m(t) = \sum_n a_n \operatorname{senc}\left(\frac{t}{T} - n\right)$$

donde $1/T$ es la tasa de transmisión de señales en bauds.

b) Los niveles de amplitud son $a_n = \pm A/2, \pm 3A/2, \ldots, \pm(M-1)\,A/2$ si M es par, y $a_n = 0, \pm A, \ldots, \pm(M-1)A/2$ si M es impar.

c) Los M niveles son de igual probabilidad y los símbolos que se transmiten en periodos adyacentes son estadísticamente independientes.

d) El ruido del canal $w(t)$ es blanco y gaussiano con media cero y densidad espectral de potencia $N_0/2$.

e) El filtro pasobajas es ideal con ancho de banda $B = 1/2T$.

f) Los niveles de umbral que se usan en el dispositivo de decisión son $0, \pm A, \ldots, \pm(M-2)A/2$ si M es par, y $\pm A/2, \pm 3A/2, \ldots, \pm(M-2)A/2$ si M es impar.

La probabilidad del error de símbolo en este sistema está definida por

$$P_e = \left(1 - \frac{1}{M}\right) \operatorname{ferc}\left(\frac{A}{2\sqrt{2\sigma}}\right)$$

donde σ es la desviación estándar del ruido a la entrada del dispositivo de decisión. Demuestre la validez de esta fórmula general determinando P_e para los siguientes tres casos: M = 2, 3, 4.

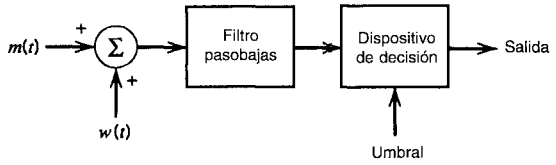


FIGURA P4.27

4.28 Suponga que en el sistema PAM M-ario de banda base con M niveles de amplitud igualmente probables, según se describe en el problema 4.27, la probabilidad promedio del error de símbolo P_e es menor que 10^{-6}, de modo que es posible ignorar la ocurrencia de errores de codificación. Demuestre que el valor mínimo de la relación señal a ruido recibida en un sistema de este tipo está dada aproximadamente por

$$(\text{SNR})_{\text{mín}} \simeq 7.8(M^2 - 1)$$

Líneas digitales de suscriptor

4.29 La distribución de amplitud de la diafonía en una línea digital de suscriptor puede modelarse como gaussiana. Justifique la validez de un modelo de tal tipo. *Sugerencia:* generalmente, un cable contiene muchos pares trenzados.

4.30 a) Obtenga la fórmula para la densidad espectral de potencia de una señal transmitida utilizando el código de línea 2B1Q.

b) Grafique el espectro de potencia de los siguientes códigos de línea:

- Código Manchester
- Código duobinario modificado
- Código bipolar con retorno a cero
- Código 2B1Q

Por tanto, compare los méritos relativos de estos códigos de línea con respecto a su conveniencia en un lazo digital de suscriptor.

4.31 En este problema utilizamos el algoritmo de mínimos cuadrados en la formulación de un cancelador de eco adaptable para utilizarlo en una línea digital de suscriptor. El principio básico de la cancelación de eco adaptable corresponde a sintetizar una réplica del eco y restarla de la señal de regreso de una manera adaptable, como se ilustra en la figura P4.31. El eco sintetizado, denotado por $\hat{r}[n]$, se genera al pasar la señal transmitida por un filtro adaptable que acopla idealmente la función de transferencia de la trayectoria del eco. La señal de retorno, consistente en la suma del eco real $r[n]$ y de la señal recibida $x[n]$, puede considerarse como la respuesta deseada en el proceso de filtrado adaptable.

Utilizando el algoritmo de mínimos cuadrados, formule la ecuación que define la operación del cancelador de eco adaptable de la figura P4.31.

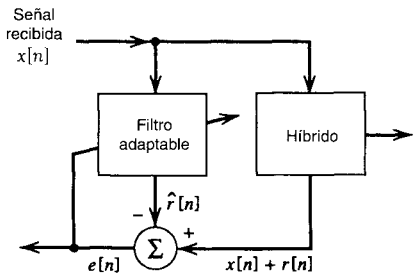


FIGURA P4.31

Ecualización

4.32 La figura P4.32 muestra la conexión en cascada de un canal lineal y de un ecualizador de retardo de línea en derivación; la respuesta al impulso del canal se denota por medio de $c(t)$ y la del ecualizador mediante $h(t)$. La $h(t)$ se define a partir de

$$h(t) = \sum_{k=-N}^{N} w_k \, \delta(t - kT)$$

donde T es el espaciamiento entre derivaciones adyacentes del ecualizador, y las w_k son los pesos (coeficientes) de derivación. La respuesta al impulso del sistema en cascada de la figura P4.32 se denota como $p(t)$. La $p(t)$ se muestrea de manera uniforme a una velocidad $1/T$. Para eliminar la diferencia de intersímbolo requerimos que se cumpla el criterio de Nyquist para la transmisión sin distorsión, como indica

$$p(nT) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

a) Imponiendo esta condición, demuestre que los $(2N + 1)$ pesos en derivación del ecualizador de forzamiento cero que se produce satisfacen el siguiente conjunto de $(2N + 1)$ ecuaciones simultáneas:

$$\sum_{k=-N}^{N} w_k c_{n-k} = \begin{cases} 1, & n = 0 \\ 0, & n \neq \pm 1, \pm 2, \ldots, \pm N \end{cases}$$

donde $c_n = c(nT)$. Por tanto, demuestre que el ecualizador de forzamiento cero es un *filtro inverso* en cuanto a que su función de transferencia es igual al recíproco de la función de transferencia del canal.

b) Una desventaja del ecualizador de forzamiento cero es el *incremento del ruido* que puede originar un pobre desempeño en la presencia de ruido del canal. Para explorar este fenómeno, considere un canal pasobajas con una muesca a la frecuencia de Nyquist, esto es, $H(f)$ es cero a $f = 1/2T$. Suponiendo que el ruido del canal es aditivo y blanco, demuestre que la densidad espectral de potencia del ruido a la salida del ecualizador tiende a infinito en $f = 1/2T$.

Incluso si el canal no tiene muesca en su respuesta en frecuencia, la densidad espectral de potencia del ruido a la salida del ecualizador puede tomar valores altos. Justifique la validez de este enunciado general.

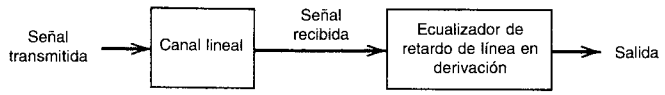


FIGURA P4.32

4.33 Considere la ecuación (4.108), la cual define la respuesta al impulso de un receptor de error cuadrático medio.

a) Justifique la validez de la ecuación (4.109), que es una versión de la transformada de Fourier de la ecuación (4.108).

b) La densidad espectral de potencia $S_q(f)$ en la ecuación (4.109) es la transformada de Fourier de la autocorrelación $R_q(\tau_1, \tau_2)$ de la función de tiempo $q(t)$. La $R_q(\tau_1, \tau_2)$ está definida por la ecuación (4.98). Partiendo de esta ecuación, deduzca la fórmula de la ecuación (4.111).

4.34 Algunos sistemas de radio sufren de *distorsión de multitrayectoria*, la cual es el resultado de la existencia de más de una trayectoria de propagación entre el transmisor y el receptor. Considere un canal cuya salida, en respuesta a la señal $s(t)$, está definida (a falta de ruido) por

$$x(t) = a_1 s(t - t_{01}) + a_2 s(t - t_{02})$$

donde a_1 y a_2 son constantes, y t_{01} y t_{02} representan retardos de la transmisión. Se propone utilizar el filtro de retardo de línea de tres derivaciones de la figura P4.34 para ecualizar la distorsión de multitrayectoria producida por este canal.

a) Evalúe la función de transferencia del canal.

b) Evalúe los parámetros del filtro de retardo de línea en derivación en términos de a_1, a_2, t_{01} y t_{02}, suponiendo que $a_2 << a_1$ y $t_{02} > t_{01}$.

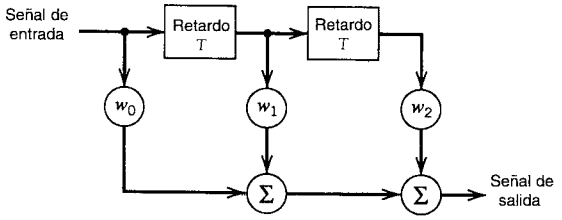


FIGURA P4.34

4.35 Considere que la secuencia $\{x(nT)\}$ denota la entrada que se aplica a un ecualizador de retardo de línea en derivación. Demuestre que el ecualizador elimina por completo la interferencia de intersímbolo siempre que su respuesta en frecuencia satisfaga la condición

$$H(f) = \frac{T}{\sum_h X(f - k/T)}$$

donde T es la duración del símbolo.

Cuando el número de derivaciones en el ecualizador tiene infinito, su respuesta en frecuencia se convierte en una serie de Fourier con coeficientes reales y, por tanto, es posible aproximar cualquier función en el intervalo $(-1/2T, 1/2T)$. Demuestre esta propiedad del ecualizador.

4.36 El parámetro de tamaño del escalón μ desempeña un papel crítico en la operación del algoritmo de mínimos cuadrados. En este contexto, analice los siguientes dos aspectos:

a) *Estabilidad.* Si μ sobrepasa cierto valor crítico, el algoritmo diverge (es decir, el sistema se vuelve inestable).

b) *Memoria.* Es posible considerar el recíproco de μ como una medida de la memoria del algoritmo: cuando hacemos μ más pequeño, una mayor cantidad de muestras pasadas de la señal de entrada influyen en la operación del algoritmo.

4.37 Sean los vectores $\mathbf{w}^{(1)}[n]$ y $\mathbf{w}^{(2)}[n]$ los pesos en derivación de las secciones prealimentada y realimentada del ecualizador de decisión realimentado de la figura 4.32. Formule el algoritmo de mínimos cuadrados para ajustar los pesos en derivación de este ecualizador.

Experimentos de computadora

4.38 En la sección 4.11 estudiamos el diagrama del ojo de un sistema de transmisión de base PAM cuaternario (M = 4) bajo condiciones tanto ruidosa como limitada en banda. En ese experimento se supuso que el canal era lineal. En un sistema estrictamente lineal con datos en verdad aleatorios, todas las aperturas del ojo serían idénticas. Sin embargo, en la práctica muchas veces es posible distinguir las simetrías en el patrón del ojo, las cuales resultan de las no linealidades en el canal de comunicación.

En este experimento estudiamos el efecto de un canal no lineal en las aperturas de un patrón del ojo. Específicamente, repetimos los experimentos de computadora correspondientes al patrón del ojo sin ruido de la figura 4.34*a* para M = 4, aunque esta vez se supone que el canal es no lineal con la siguiente relación entrada-salida:

$$x(t) = s(t) + as^2(t)$$

donde $s(t)$ es la entrada del canal y $x(t)$ es la salida del mismo, y a es una constante.

a) Realice el experimento para $a = 0, 0.05, 0.1, 0.2$.

b) En consecuencia, analice cómo la a variable afecta la forma del patrón del ojo.

4.39 En este experimento estudiamos el *pulso de la raíz del coseno elevado* debido a Chennakeshu y Saulnier. Este pulso, denotado por $p(t)$, tiene las siguientes propiedades:

- El pulso $p(t)$ es simétrico en el tiempo, esto es, $p(-t) = p(t)$.
- La transformada de Fourier al cuadrado de $p(t)$, a saber, $P^2(f)$, satisface el espectro del coseno derivado de la ecuación (4.60), aunque no ocurre lo mismo con la propia transformada de Fourier $P(f)$.
- El pulso $p(t)$ satisface la *restricción de ortogonalidad*:

$$\int_{-\infty}^{\infty} p(t)p(t-nT)\,dt = 0, \quad n = \pm 1, \pm 2, \ldots$$

donde T es el periodo del símbolo.

a) Calcule la forma de onda de la banda base de la secuencia de datos binaria 101100 para un factor de rodado $\alpha = 0.3$.

b) Compare la forma de onda computada en la parte *a* con la que se obtiene utilizando el espectro de coseno elevado ordinario.

CAPÍTULO 5

ANÁLISIS ESPACIAL DE SEÑALES

Este capítulo expone algunos aspectos básicos que pertenecen a la transmisión de señales sobre un canal de ruido blanco gaussiano aditivo (AWGN). De manera específica, aborda los siguientes temas:

- *La representación geométrica de señales con energía finita, la cual proporciona una herramienta elegantemente matemática y de gran profundidad para el estudio de la transmisión de datos.*
- *El procedimiento de máxima verosimilitud para la detección de una señal en un canal AWGN.*
- *La deducción del receptor de correlación que es equivalente al receptor de filtro de acoplamiento que se explicó en el capítulo anterior.*
- *La probabilidad del error de símbolo y de la cota de unión para su cálculo aproximado.*

El material que se presenta aquí conduce en forma natural al estudio de la transmisión de datos pasobanda que se aborda en el capítulo 6.

5.1 *Introducción*

Considere la forma más básica de un sistema de comunicación digital descrito en la figura 5.1. Una *fuente del mensaje* emite un *símbolo* cada T segundos, con los símbolos pertenecientes a un alfabeto de M símbolos denotados por $m_1, m_2, \ldots, m_M$. Considere, por ejemplo, la conexión remota de dos computadoras digitales, con una computadora actuando como una fuente de información que calcula salidas digitales basadas en observaciones y entradas que se le alimentan. La salida de computadora que se produce se expresa como una secuencia de ceros y unos, que se transmiten a una segunda computadora por un canal de comunicación. En este caso, el alfabeto consta simplemente de dos símbolos binarios: 0 y 1. Un segundo ejemplo es el de un codificador PCM cuaternario con un alfabeto compuesto por cuatro símbolos posibles: 00, 01, 10 y 11. En cualquier caso, las probabilidades *a priori* $p_1, p_2, \ldots, p_M$ especifican la salida de la fuente del mensaje. Si se carece de información previa, suele suponerse que los M símbolos del alfabeto son *igualmente probables*. Es posible expresar entonces la probabilidad de que m_i símbolos sean emitidos por la fuente como

$$\begin{aligned} p_i &= P(m_i) \\ &= \frac{1}{M} \text{ para } i = 1, 2, \ldots, M \end{aligned} \tag{5.1}$$

El transmisor toma la salida m_i de la fuente del mensaje y la codifica en una señal *distinta* $s_i(t)$ adecuada para la transmisión por el canal. La señal $s_i(t)$ ocupa la duración completa T asignada al símbolo m_i. Es más importante el hecho de que $s_i(t)$ es una *señal de energía* de valores reales (es decir, una señal con energía finita), como se indica por medio de

$$E_i = \int_0^T s_i^2(t)\, dt, \quad i = 1, 2, \ldots, M \tag{5.2}$$

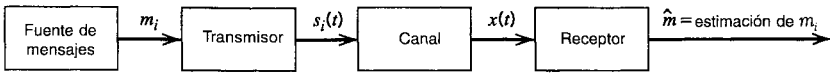


FIGURA 5.1 Diagrama de bloques de un sistema de comunicación digital general.

Se supone que el canal tiene dos características:

1. El canal es *lineal*, con un ancho de banda lo suficientemente amplio para permitir la transmisión de la señal $s_i(t)$ sin o con una distorsión despreciable.
2. El ruido del canal, $w(t)$, es una función de muestreo de un *proceso de ruido blanco gaussiano de media cero*. Las razones para esta segunda suposición es porque permite los cálculos relativos al receptor y constituye una descripción razonable del tipo de ruido presente en muchos sistemas de comunicación prácticos.

Nos referimos a un canal de este tipo como un *canal del ruido blanco gaussiano aditivo* (AWGN). En consecuencia, es posible expresar la *señal recibida* $x(t)$ como

$$x(t) = s_i(t) + w(t), \quad \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.3}$$

y de esa manera modelar el canal como en la figura 5.2.

El receptor tiene la tarea de observar la señal recibida $x(t)$ para una duración de T segundos y efectuar una mejor *estimación* de la señal transmitida $s_i(t)$ o, equivalentemente, del símbolo m_i. Sin embargo, por la presencia del ruido del canal, este proceso de toma de decisiones es de naturaleza estadística, resultando que el receptor efectuará errores ocasionales. Entonces se requiere diseñar el receptor de manera que minimice la *probabilidad promedio de errores de símbolo*, que se define como

$$P_e = \sum_{i=1}^{M} p_i \, P(\hat{m} \neq m_i | m_i) \tag{5.4}$$

donde m_i es el símbolo transmitido, $\hat{m}$ es la estimación producida por el receptor y $P(\hat{m} \neq m_i | m_i)$ es la probabilidad de error condicional, dado que se envió el i-ésimo símbolo. Se afirma que el receptor que se produce será *óptimo en el sentido de la probabilidad mínima de error*.

Este modelo ofrece las bases para el diseño del receptor óptimo, para el cual usaremos la representación geométrica del conjunto conocido de señales transmitidas, $\{s_i(t)\}$. Este método, que se explica en la sección 5.2, proporciona una gran cantidad de información, con una simplificación considerable de los detalles.

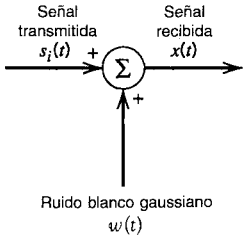


FIGURA 5.2 Modelo del ruido blanco gaussiano aditivo (AWGN) de un canal.

5.2 Representación geométrica de señales

La esencia de la *representación geométrica de señales*[1] radica en representar cualquier conjunto de M señales de energía $\{s_i(t)\}$ como combinaciones lineales de N *funciones de base ortonormal*, donde $N \leq M$. Es decir, dado un conjunto de señales de energía de valor real $s_1(t), s_2(t), \ldots, s_M(t)$, cada una de T segundos de duración, escribir

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t), \quad \begin{cases} 0 \leq t \leq T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.5}$$

donde los coeficientes de la expansión están definidos por

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt, \begin{cases} i = 1, 2, \ldots, M \\ j = 1, 2, \ldots, N \end{cases} \tag{5.6}$$

Las funciones básicas de valor real $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ son *ortonormales*, lo cual quiere decir que

$$\int_0^T \phi_i(t)\phi_j(t)\,dt = \delta_{ij} = \begin{cases} 1 \text{ si } i = j \\ 0 \text{ si } i \neq j \end{cases} \tag{5.7}$$

donde δ_{ij} es la *delta de Kronecker*. La primera condición de la ecuación (5.7) establece que cada función básica se *normaliza* para tener energía unitaria. La segunda condición plantea que las funciones básicas $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ son *ortogonales* entre sí con respecto al intervalo $0 \leq t \leq T$.

El conjunto de coeficientes $\{s_{ij}\}_{j=1}^{N}$ puede considerarse de modo natural como un *vector N-dimensional* denotado por $\mathbf{s}_i$. El punto importante por advertir aquí es que el vector $\mathbf{s}_i$ mantiene una relación *uno a uno* con la señal transmitida $s_i(t)$:

- Dados los N elementos de los vectores $\mathbf{s}_i$ (es decir, $s_{i1}, s_{i2}, \ldots, s_{iN}$) operando como entrada, es posible utilizar el esquema que se muestra en la figura 5.3*a* para generar la señal $s_i(t)$, la

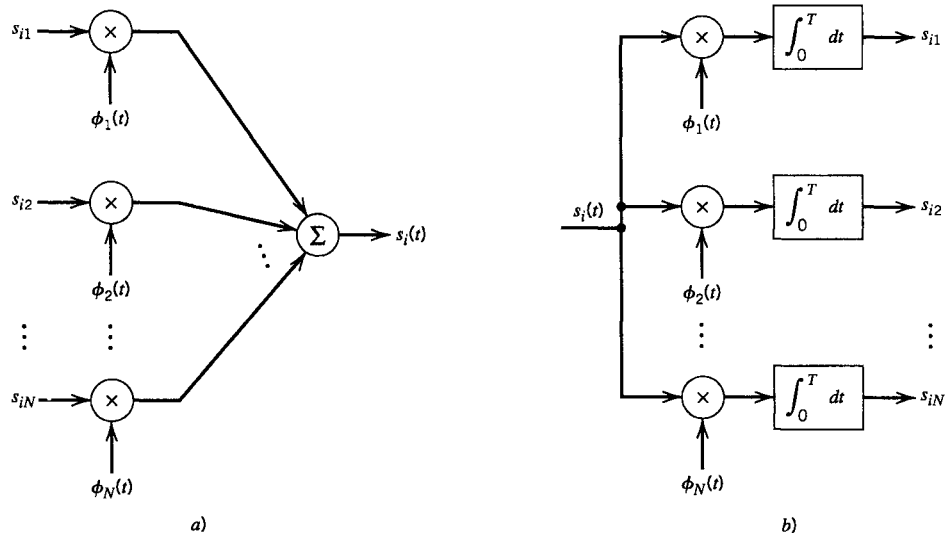


FIGURA 5.3 *a*) Sintetizador para generar la señal $s_i(t)$. *b*) Analizador para generar el conjunto de vectores de señal $\{s_i\}$.

cual se deriva directamente de la ecuación 5.5. Ésta consiste en un banco de N multiplicadores, cada uno con su propia función básica, seguido por un sumador. Este esquema puede considerarse como un *sintetizador.*

- Contrariamente, dadas las señales $s_i(t)$, $i = 1, 2, \ldots, M$, que operan como entrada, resulta viable utilizar el esquema que se muestra en la figura 5.3*b* para calcular los coeficientes s_{i1}, $s_{i2}, \ldots, s_{iN}$ que surgen directamente de la ecuación (5.6). Este segundo esquema consiste en un banco de N *integradores de producto* o *correlatores*, con una entrada común, y con cada uno de ellos alimentados con su propia función básica. El esquema de la figura 5.3*b* puede verse como un *analizador.*

Por tanto, es posible establecer que cada señal en el conjunto $\{s_i(t)\}$ está determinada por completo por el *vector* a partir de sus coeficientes

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, \qquad i = 1, 2, \ldots, M \tag{5.8}$$

El vector $\mathbf{s}_i$ recibe el nombre de *vector de señal.* Asimismo, si ampliamos conceptualmente nuestra noción convencional de espacios euclidianos de dos y tres dimensiones a un *espacio euclidiano N-dimensional*, podemos visualizar el conjunto de vectores de señal $\{\mathbf{s}_i | i = 1, 2, \ldots, M\}$ como el que define un conjunto correspondiente de M puntos en un espacio euclidiano N-dimensional, con N ejes mutuamente perpendiculares denominados $\phi_1, \phi_2, \ldots, \phi_N$. Este espacio euclidiano N-dimensional recibe el nombre de *espacio de la señal.*

La idea de visualizar geométricamente un conjunto de señales de energía, como acabamos de describir, es de una profunda importancia. Proporciona la base matemática para la representación geométrica de las señales de energía, abriendo el camino para el análisis de ruido de los sistemas de comunicación digitales de una manera conceptualmente satisfactoria. Esta forma de representación se ilustra en la figura 5.4 para el caso de un espacio de señal bidimensional con tres señales, esto es, $N = 2$ y $M = 3$.

En un espacio euclidiano N-dimensional podemos definir *longitudes* de vectores y *ángulos* entre vectores. Suele denotarse la longitud (denominada también el *valor absoluto* o *norma)* de un vector de señal $\mathbf{s}_i$ mediante el símbolo $\|\mathbf{s}_i\|$. La longitud al cuadrado de cualquier vector de señal $\mathbf{s}_i$ se define como el *producto interno* o *producto punto* de $\mathbf{s}_i$ consigo misma, como se muestra por medio de

$$\begin{aligned} \|\mathbf{s}_i\|^2 &= \mathbf{s}_i^T \mathbf{s}_i \\ &= \sum_{j=1}^{N} s_{ij}^2, \quad i = 1, 2, \ldots, M \end{aligned} \tag{5.9}$$

donde s_{ij} es el j-ésimo elemento de $\mathbf{s}_i$ y T denota la transposición de matriz.

Existe una relación interesante entre el contenido de energía de una señal y su representación como un vector. Por definición, la energía de una señal $s_i(t)$ de T segundos de duración es

$$E_i = \int_0^T s_i^2(t)\, dt \tag{5.10}$$

Por tanto, al sustituir la ecuación (5.5) en la (5.10), obtenemos

$$E_i = \int_0^T \left[\sum_{j=1}^{N} s_{ij} \phi_j(t) \right] \left[\sum_{k=1}^{N} s_{ik} \phi_k(t) \right] dt$$

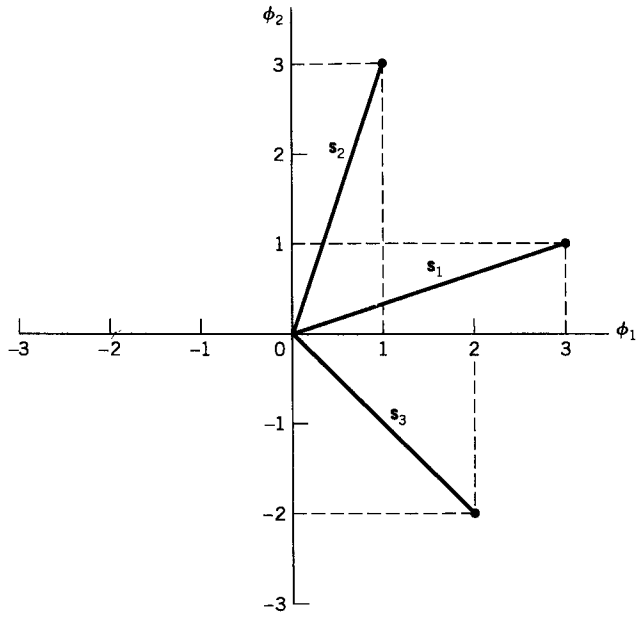


FIGURA 5.4 Ilustración de la representación geométrica de señales para el caso en el que $N = 2$ y $M = 3$.

Intercambiando el orden de la sumatoria y la integración, y rearreglando después los términos, obtenemos

$$E_i = \sum_{j=1}^{N} \sum_{k=1}^{N} s_{ij} s_{ik} \int_0^T \phi_j(t)\phi_k(t)\, dt \tag{5.11}$$

Sin embargo, puesto que las $\phi_j(t)$ forman un conjunto ortonormal, de acuerdo con las dos condiciones de la ecuación (5.7), encontramos que la ecuación (5.11) se reduce simplemente a

$$\begin{aligned} E_i &= \sum_{j=1}^{N} s_{ij}^2 \\ &= \|\mathbf{s}_i\|^2 \end{aligned} \tag{5.12}$$

De ese modo, las ecuaciones (5.9) y (5.12) muestran que la energía de una señal $s_i(t)$ es igual a la longitud del vector de señal $\mathbf{s}_i(t)$ que la representa.

En el caso de una par de señales $s_i(t)$ y $s_k(t)$, representadas por los vectores de señal $\mathbf{s}_i$ y $\mathbf{s}_k$, respectivamente, podemos demostrar también que

$$\int_0^T s_i(t)s_k(t)\, dt = \mathbf{s}_i^T \mathbf{s}_k \tag{5.13}$$

La ecuación (5.13) establece que el *producto interno* de las señales $s_i(t)$ y $s_k(t)$ sobre el intervalo $[0, T]$, utilizando sus representaciones en el dominio del tiempo, es igual al producto interno de sus representaciones de vector respectivas $\mathbf{s}_i$ y $\mathbf{s}_k$. Advierta que el producto interno de $s_i(t)$ y $s_k(t)$ es *invariante* para la elección de las funciones básicas $\{\phi_j(t)\}_{j=1}^N$ con respecto a que depende únicamente de las componentes de las señales $s_i(t)$ y $s_k(t)$ proyectadas sobre cada una de las funciones básicas.

Otra relación aún útil que implica las representaciones de vector de las señales $s_i(t)$ y $s_k(t)$ se describe mediante

$$\begin{aligned}\|\mathbf{s}_i - \mathbf{s}_k\|^2 &= \sum_{j=1}^{N}(s_{ij} - s_{kj})^2 \\ &= \int_0^T (s_i(t) - s_k(t))^2\,dt\end{aligned} \tag{5.14}$$

donde $\|\mathbf{s}_i - \mathbf{s}_k\|$ es la *distancia euclidiana*, d_{ik}, entre los puntos representados por los vectores de señal $\mathbf{s}_i$ y $\mathbf{s}_k$.

Para completar la representación geométrica de las señales de energía, necesitamos contar con una representación para el ángulo θ_{ik} subtendido entre dos vectores de señal $\mathbf{s}_i$ y $\mathbf{s}_k$. Por definición, el *coseno del ángulo* θ_{ik} es igual al producto interno de estos dos vectores dividido por el producto de sus normas individuales,

$$\cos\theta_{ik} = \frac{\mathbf{s}_i^T\mathbf{s}_k}{\|\mathbf{s}_i\|\,\|\mathbf{s}_k\|} \tag{5.15}$$

Los dos vectores $\mathbf{s}_i$ y $\mathbf{s}_k$ son de tal modo *ortogonales* o *perpendiculares* entre sí si su producto interno $\mathbf{s}_i^T\mathbf{s}_k$ es cero, $\theta_{ik} = 90$ grados; esta condición se satisface intuitivamente.

▶ EJEMPLO 5.1 Desigualdad de Schwarz

Considere cualquier par de señales de energía $s_1(t)$ y $s_2(t)$. La *desigualdad de Schwarz* establece que

$$\left(\int_{-\infty}^{\infty} s_1(t)s_2(t)dt\right)^2 \le \left(\int_{-\infty}^{\infty} s_1^2(t)dt\right)\left(\int_{-\infty}^{\infty} s_2^2(t)dt\right) \tag{5.16}$$

La desigualdad se cumple si y sólo si $s_2(t) = cs_1(t)$, donde c es cualquier constante.

Para demostrar esta importante desigualdad, dejemos que $s_1(t)$ y $s_2(t)$ se expresen en términos del par de funciones básicas ortogonales $\phi_1(t)$ y $\phi_2(t)$ del modo siguiente:

$$\begin{aligned}s_1(t) &= s_{11}\phi_1(t) + s_{12}\phi_2(t) \\ s_2(t) &= s_{21}\phi_1(t) + s_{22}\phi_2(t)\end{aligned}$$

donde $\phi_1(t)$ y $\phi_2(t)$ satisface las condiciones de ortogonalidad sobre el intervalo de tiempo completo $(-\infty, \infty)$:

$$\int_{-\infty}^{\infty} \phi_i(t)\phi_j(t)dt = \delta_{ij} = \begin{cases}1 & \text{para } j = i \\ 0 & \text{en otro caso}\end{cases}$$

Sobre esta base, es posible representar las señales $s_1(t)$ y $s_2(t)$ mediante el siguiente par respectivo de vectores, como se ilustra en la figura 5.5:

$$\mathbf{s}_1 = \begin{bmatrix}s_{11} \\ s_{12}\end{bmatrix}$$

$$\mathbf{s}_2 = \begin{bmatrix}s_{21} \\ s_{22}\end{bmatrix}$$

En la figura 5.5 vemos que el ángulo θ subtendido entre los vectores $\mathbf{s}_1$ y $\mathbf{s}_2$ es

$$\begin{aligned}\cos\theta &= \frac{\mathbf{s}_1^T\mathbf{s}_2}{\|\mathbf{s}_1\|\,\|\mathbf{s}_2\|} \\ &= \frac{\int_{-\infty}^{\infty} s_1(t)s_2(t)dt}{\left(\int_{-\infty}^{\infty} s_1^2(t)dt\right)^{1/2}\left(\int_{-\infty}^{\infty} s_2^2(t)dt\right)^{1/2}}\end{aligned} \tag{5.17}$$

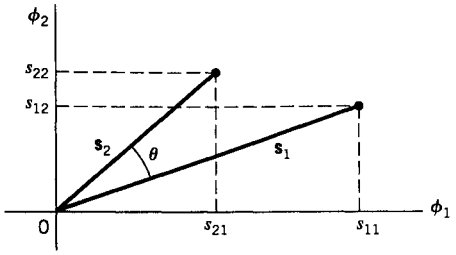


FIGURA 5.5 Representación vectorial de las señales $s_1(t)$ y $s_2(t)$, proporcionándose la imagen de fondo para demostrar la desigualdad de Schwarz.

donde hemos recurrido a las ecuaciones (5.15), (5.13) y (5.9). Admitiendo que $|\cos\theta| \leq 1$, la desigualdad de Schwarz de la ecuación (5.16) se desprende de inmediato de la ecuación (5.17). Asimismo, en la primera línea de la ecuación (5.17) advertimos que $|\cos\theta| = 1$ si y sólo si $\mathbf{s}_2 = c\mathbf{s}_1$, esto es, $s_2(t) = cs_1(t)$, donde c es una constante arbitraria.

La demostración de la desigualdad de Schwarz, de acuerdo como se presenta aquí, se aplica a señales de valores reales. Es posible extenderla sin dificultades a señales de valores complejos, en cuyo caso la ecuación (5.16) se reformula como

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)dt\right|^2 \leq \left(\int_{-\infty}^{\infty}|s_1(t)|^2 dt\right)\left(\int_{-\infty}^{\infty}|s_2(t)|^2 dt\right) \tag{5.18}$$

donde la desigualdad se cumple si y sólo si $s_2(t) = cs_1(t)$, donde c es una constante; vea el problema 5.9. La forma compleja de la desigualdad de Schwarz fue la que se usó en el capítulo 4 para obtener el filtro de acoplamiento. ◀

■ PROCEDIMIENTO DE ORTOGONALIZACIÓN DE GRAM-SCHMIDT

Luego de demostrar la elegancia de la representación geométrica de las señales de energía, ¿cómo la justificamos en términos matemáticos? La respuesta se ubica en el *procedimiento de ortogonalización de Gram-Schmidt*, para el cual necesitamos un *conjunto ortonormal completo de funciones básicas*. Para proseguir con la formulación de este procedimiento, suponga que tenemos un conjunto de M señales de energía denotado por $s_1(t), s_2(t), \ldots, s_M(t)$. Partiendo de $s_1(t)$ elegido arbitrariamente de este conjunto, la primera función básica está definida por

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}} \tag{5.19}$$

donde E_1 es la energía de la señal $s_1(t)$. En ese caso, claramente tenemos

$$\begin{aligned} s_1(t) &= \sqrt{E_1}\phi_1(t) \\ &= s_{11}\phi_1(t) \end{aligned} \tag{5.20}$$

donde los coeficientes $s_{11} = \sqrt{E_1}$ y $\phi_1(t)$ tienen energía unitaria, como se requiere.

A continuación, utilizando la señal $s_2(t)$, definimos el coeficiente s_{21} como

$$s_{21} = \int_0^T s_2(t)\phi_1(t)dt \tag{5.21}$$

Es posible de ese modo introducir una nueva función intermedia

$$g_2(t) = s_2(t) - s_{21}\phi_1(t) \tag{5.22}$$

que es ortogonal a $\phi_1(t)$ sobre el intervalo $0 \le t \le T$ en virtud de la ecuación (5.21) y el hecho de que la función básica $\phi_1(t)$ tiene energía unitaria. Después de esto estamos listos para definir la segunda función básica como

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)dt}} \tag{5.23}$$

Al sustituir la ecuación (5.22) en la (5.23) y simplificar, obtenemos el resultado que se desea

$$\phi_2(t) = \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{E_2 - s_{21}^2}} \tag{5.24}$$

donde E_2 es la energía de la señal $s_2(t)$. Resulta claro de la ecuación (5.23) que

$$\int_0^T \phi_2^2(t)dt = 1$$

y a partir de la ecuación (5.24) que

$$\int_0^T \phi_1(t)\phi_2(t)dt = 0$$

En otras palabras, $\phi_1(t)$ y $\phi_2(t)$ forman un par ortogonal, como se requiere.

Continuando de esta forma, podemos definir en general

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t) \tag{5.25}$$

donde los coeficientes s_{ij} se definen a sí mismos por medio de

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)dt, \quad j = 1, 2, \ldots, i-1 \tag{5.26}$$

La ecuación (5.22) es un caso especial de la (5.25) con $i = 2$. Advierta también que para $i = 1$, la función $g_i(t)$ se reduce a $s_i(t)$.

Dada la $g_i(t)$, podemos definir ahora un conjunto de funciones básicas

$$\phi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)dt}}, \quad i = 1, 2, \ldots, N \tag{5.27}$$

que constituyen un conjunto ortonormal. La dimensión N es menor que o igual que el número de señales dadas, M, dependiendo de una de dos posibilidades:

- Las señales $s_1(t), s_2(t), \ldots, s_M(t)$ forman un *conjunto linealmente independiente*, en cuyo caso $N = M$.
- Las señales $s_1(t), s_2(t), \ldots, s_M(t)$ *no* son linealmente independientes, en cuyo caso $N < M$, y la función intermedia $g_i(t)$ es cero para $i > N$.

TABLA 5.1 *Niveles de amplitud del código 2B1Q*

Señal		
Símbolo	*Amplitud*	*Código de Gray*
$s_1(t)$	−3	00
$s_2(t)$	−1	01
$s_3(t)$	+1	11
$s_4(t)$	+3	10

Advierta que la expansión convencional en series de Fourier de una señal periódica es un ejemplo de una expansión particular del tipo que se describe aquí. Además, la representación de una señal limitada en banda en términos de sus muestras tomadas a la frecuencia de Nyquist puede considerarse como otra muestra de una expansión particular de este tipo. Sin embargo, resulta necesario hacer dos distinciones:

1. La forma de las funciones básicas $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ no se ha especificado. Es decir, a diferencia de la expansión en series de Fourier de una señal periódica o de la representación muestreada de una señal limitada en banda, no hemos restringido el procedimiento de ortogonalización de Gram-Schmidt para que esté en términos de las funciones senoidales o las funciones senc de tiempo.
2. La expansión de la señal $s_i(t)$ en términos de un número finito de términos no es una aproximación donde únicamente los primeros N términos sean significativos, sino que más bien es una *expresión exacta* donde N y sólo N términos son significativos.

▶ EJEMPLO 5.2 Código 2B1Q

El código 2B1Q se describió en el capítulo 4 como el código de línea estadounidense para líneas digitales de suscriptores. Éste representa una señal PAM cuaternaria, como se muestra en el alfabeto de codificación Gray de la tabla 5.1. Las cuatro señales posibles, $s_1(t)$, $s_2(t)$, $s_3(t)$ y $s_4(t)$, son versiones escaladas en amplitud de un pulso de Nyquist. Cada señal representa un dibit. Deseamos determinar la representación vectorial del código 2B1Q.

Este ejemplo es lo suficientemente simple para que lo resolvamos por inspección. Dejemos que $\phi_1(t)$ denote el pulso de Nyquist, normalizado para que tenga energía unitaria. La $\phi_1(t)$ definida de esa forma es la única función básica para la representación vectorial del código 2B1Q. En consecuencia, la representación del espacio de la señal de este código es como se ilustra en la figura 5.6. Ésta consiste en cuatro vectores de señal, $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ y $\mathbf{s}_4$, que se ubican sobre el eje ϕ_1 de manera simétrica alrededor del origen. En este ejemplo tenemos, de tal modo, $M = 4$ y $N = 1$.

Es posible generalizar el resultado descrito en la figura 5.6 para el código 2B1Q como sigue. El diagrama del espacio de la señal de una señal modulada por amplitud de pulso M-aria, en general, es unidimensional con M puntos de señal posicionados de manera uniforme sobre el único eje del diagrama. ◀

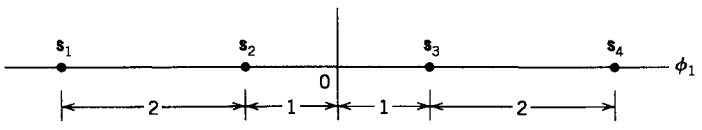


FIGURA 5.6 Representación del espacio de la señal del código 2B1Q.

5.3 Conversión del canal AWGN continuo en un canal vectorial

Suponga que la entrada al banco de N integradores de producto o correlatores en la figura 5.3*b* no es la señal transmitida $s_i(t)$, sino más bien la señal recibida $x(t)$ definida de acuerdo con el canal AWGN idealizado de la figura 5.2. En otras palabras,

$$x(t) = s_i(t) + w(t), \quad \begin{cases} 0 \le t \le T \\ i = 1, 2, \ldots, M \end{cases} \tag{5.28}$$

donde $w(t)$ es una función de muestreo de un proceso de ruido blanco gaussiano $W(t)$ de media cero y densidad espectral de potencia $N_0/2$. En forma correspondiente, encontramos que la salida del correlator j, digamos, es el valor de muestreo de una variable aleatoria X_j, como indica

$$\begin{aligned} x_j &= \int_0^T x(t)\phi_j(t)dt \\ &= s_{ij} + w_j, \qquad j = 1, 2, \ldots, N \end{aligned} \tag{5.29}$$

La primera componente, s_{ij}, es una cantidad determinística que aporta la señal transmitida $s_i(t)$; está definida por

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)dt \tag{5.30}$$

La segunda componente, w_j, es el valor de muestreo de una variable aleatoria W_j que surge debido a la presencia del ruido $w(t)$ del canal; se define como

$$w_j = \int_0^T w(t)\phi_j(t)dt \tag{5.31}$$

Considere a continuación un nuevo proceso aleatorio $X'(t)$ cuya función de muestreo $x'(t)$ se relaciona con la señal recibida $x(t)$ del siguiente modo:

$$x'(t) = x(t) - \sum_{j=1}^{N} x_j\phi_j(t) \tag{5.32}$$

La sustitución de las ecuaciones (5.28) y (5.29) en la (5.32), y el empleo posterior de la expansión de la ecuación (5.5), origina

$$\begin{aligned} x'(t) &= s_i(t) + w(t) - \sum_{j=1}^{N} (s_{ij} + w_j)\phi_j(t) \\ &= w(t) - \sum_{j=1}^{N} w_j\phi_j(t) \\ &= w'(t) \end{aligned} \tag{5.33}$$

La función de muestreo $x'(t)$ depende, en consecuencia, exclusivamente del ruido $w(t)$ del canal. Con base en las ecuaciones (5.32) y (5.33) podemos expresar entonces la señal recibida como

$$\begin{aligned} x(t) &= \sum_{j=1}^{N} x_j\phi_j(t) + x'(t) \\ &= \sum_{j=1}^{N} x_j\phi_j(t) + w'(t) \end{aligned} \tag{5.34}$$

En consecuencia, es posible considerar $w'(t)$ como una especie de término *remanente* que debe incluirse a la derecha para preservar la igualdad en la ecuación (5.34). Resulta informativo

comparar la expansión de la señal recibida $x(t)$ dada en la ecuación (5.34) con la expansión correspondiente de la señal transmitida $s_i(t)$ indicada por la ecuación (5.5). Esta última es completamente determinística, en tanto que la de la ecuación (5.34) es aleatoria (estocástica), que era lo que se esperaba.

■ Caracterización estadística de las salidas de un correlator

Deseamos ahora desarrollar una caracterización estadística de los conjuntos de N salidas de un correlator. Dejemos que $X(t)$ denote el proceso aleatorio, una de cuyas funciones de muestreo se representa mediante la señal recibida $x(t)$. De modo correspondiente, sea X_j la variable aleatoria cuyo valor de muestreo se representa por medio de la salida del correlator x_j, $j = 1, 2, \ldots, N$. De acuerdo con el modelo AWGN de la figura 5.2, el proceso aleatorio $X(t)$ es gaussiano. En consecuencia, se concluye que X_j es una variable aleatoria gaussiana para toda j (vea la propiedad 1 de un proceso gaussiano, sección 1.8). Por tanto, X_j está caracterizado por completo por su media y varianza, las cuales se determinan a continuación.

Sea W_j la variable aleatoria representada por el valor de muestreo w_j producido por el j-ésimo correlator en respuesta a la componente de ruido blanco gaussiano $w(t)$. La variable aleatoria W_j tiene media cero, ya que el proceso de ruido $W(t)$ representado por $w(t)$ en el modelo AWGN de la figura 5.2 tiene media cero por definición. En consecuencia, la X_j depende sólo de s_{ij}, según indica

$$\begin{aligned} \mu_{X_j} &= E[X_j] \\ &= E[s_{ij} + W_j] \\ &= s_{ij} + E[W_j] \\ &= s_{ij} \end{aligned} \tag{5.35}$$

Para determinar la varianza X_j, advertimos que

$$\begin{aligned} \sigma_{X_j}^2 &= \text{var}[X_j] \\ &= E[(X_j - s_{ij})^2] \\ &= E[W_j^2] \end{aligned} \tag{5.36}$$

donde la última línea surge de la ecuación (5.29) con x_j y w_j sustituidas por X_j y W_j, respectivamente. De acuerdo con la ecuación (5.31), la variable aleatoria W_j está definida por

$$W_j = \int_0^T W(t)\phi_j(t)dt$$

Por tanto, es posible expandir la ecuación (5.36) del siguiente modo:

$$\begin{aligned} \sigma_{X_j}^2 &= E\left[\int_0^T W(t)\phi_j(t)dt \int_0^T W(u)\phi_j(u)du\right] \\ &= E\left[\int_0^T \int_0^T \phi_j(t)\phi_j(u)W(t)W(u)dtdu\right] \end{aligned} \tag{5.37}$$

Intercambiando el orden de la integración y de la esperanza:

$$\begin{aligned} \sigma_{X_j}^2 &= \int_0^T \int_0^T \phi_j(t)\phi_j(u)E[W(t)W(u)]dtdu \\ &= \int_0^T \int_0^T \phi_j(t)\phi_j(u)R_W(t,u)dtdu \end{aligned} \tag{5.38}$$

donde $R_W(t, u)$ es la función de autocorrelación del proceso de ruido $W(t)$. Puesto que este ruido es estacionario, $R_W(t, u)$ depende sólo de la diferencia de tiempo $t - u$. Además, puesto que el ruido $W(t)$ es blanco con densidad espectral de potencia constante $N_0/2$, es posible expresar $R_W(t, u)$ de la manera siguiente [vea la ecuación (1.95)]:

$$R_W(t, u) = \frac{N_0}{2}\delta(t-u) \tag{5.39}$$

Por tanto, al sustituir la ecuación (5.39) en la (5.38), y al usar después la propiedad de filtrado de la función $\delta(t)$, obtenemos

$$\begin{aligned}\sigma_{X_j}^2 &= \frac{N_0}{2}\int_0^T\int_0^T \phi_j(t)\phi_j(u)\,\delta(t-u)dtdu \\ &= \frac{N_0}{2}\int_0^T \phi_j^2(t)dt\end{aligned} \tag{5.40}$$

Puesto que la $\phi_j(t)$ tiene energía unitaria, por definición, obtenemos finalmente el resultado simple

$$\sigma_{X_j}^2 = \frac{N_0}{2} \qquad \text{para toda } j \tag{5.41}$$

Este importante resultado muestra que todas las salidas del correlator denotadas por X_j con $j = 1, 2, \ldots, N$, tienen una varianza igual a la densidad espectral de potencia $N_0/2$ del proceso de ruido $W(t)$.

Asimismo, puesto que $\phi_j(t)$ forma un conjunto ortogonal, encontramos que las X_j son mutuamente no correlacionadas, según expresa

$$\begin{aligned}\operatorname{cov}[X_j X_k] &= E[(X_j - \mu_{X_j})(X_k - \mu_{X_k})] \\ &= E[(X_j - s_{ij})(X_k - s_{ik})] \\ &= E[W_j W_k] \\ &= E\left[\int_0^T W(t)\phi_j(t)dt \int_0^T W(u)\phi_k(u)du\right] \\ &= \int_0^T\int_0^T \phi_j(t)\phi_k(u)R_W(t,u)dtdu \\ &= \frac{N_0}{2}\int_0^T\int_0^T \phi_j(t)\phi_k(u)\,\delta(t-u)dtdu \\ &= \frac{N_0}{2}\int_0^T \phi_j(t)\phi_k(t)dt \\ &= 0, \qquad j \neq k\end{aligned} \tag{5.42}$$

Puesto que las X_j son variables aleatorias gaussianas, la ecuación (5.42) implica que también son estadísticamente independientes (vea la propiedad 4 de un proceso gaussiano, sección 1.8).

Defina el vector de N variables aleatorias

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \tag{5.43}$$

cuyos elementos son variables aleatorias gaussianas e independientes con valores medios iguales a s_{ij} y varianza igual a $N_0/2$. Puesto que los elementos del vector $\mathbf{X}$ son independientes

estadísticamente, es posible expresar la función de densidad de probabilidad condicional del vector **X**, dado que se transmitió la señal $s_i(t)$ o correspondientemente el símbolo m_i, como el producto de las funciones de densidad de probabilidad condicional de sus elementos individuales:

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = \prod_{j=1}^{N} f_{X_j}(x_j|m_i), \quad i = 1, 2, \ldots, M \tag{5.44}$$

donde el vector **x** y el escalar x_j son valores de muestreo del vector aleatorio **X** y la variable X_j, respectivamente. El vector **x** recibe el nombre de *vector de observación*; de modo correspondiente, x_i se denomina *elemento observable*. Cualquier canal que satisface la ecuación (5.44) recibe el nombre de *canal sin memoria*.

Puesto que cada X_j es una variable aleatoria gaussiana con media s_{ij} y varianza $N_0/2$, tenemos

$$f_{X_j}(x_j|m_i) = \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_j - s_{ij})^2\right], \quad \begin{matrix} j = 1, 2, \ldots, N \\ i = 1, 2, \ldots, M \end{matrix} \tag{5.45}$$

Por tanto, la sustitución de la ecuación (5.45) en la (5.44) produce

$$f_{\mathbf{X}}(\mathbf{x}|m_i) = (\pi N_0)^{-N/2} \exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2\right], \quad i = 1, 2, \ldots, M \tag{5.46}$$

Ahora es claro que los elementos de un vector aleatorio **X** caracterizan completamente el término de la sumatoria $\Sigma_j X_j \phi_j(t)$, cuyo valor de muestreo se representa mediante el primer término de la ecuación (5.34). Sin embargo, permanece el término del ruido $w'(t)$ en esta ecuación, el cual depende sólo del ruido del canal $w(t)$. Puesto que el proceso de ruido $W(t)$ representado por $w(t)$ es gaussiano con media cero, se concluye que el proceso del ruido $W'(t)$ representado por la función de muestreo $w'(t)$ también es un proceso gaussiano con media cero. Por último, advertimos que cualquier variable aleatoria $W'(t_k)$, por ejemplo, derivada del proceso de ruido $W'(t)$ al muestrearla en tiempo t_k, es de hecho independiente estadísticamente del conjunto de variables aleatorias $\{X_j\}$; es decir (ver problema 5.10),

$$E[X_j W'(t_k)] = 0, \quad \begin{cases} j = 1, 2, \ldots, N \\ 0 \le t_k \le T \end{cases} \tag{5.47}$$

Puesto que cualquier variable aleatoria basada en el proceso de ruido remanente $W'(t)$ es independiente del conjunto de variables aleatorias $\{X_j\}$, así como del conjunto de señales transmitidas $\{s_i(t)\}$, la ecuación (5.47) establece que la variable aleatoria $W'(t_k)$ es irrelevante para la decisión de cuál señal particular realmente se transmite. En otras palabras, las salidas del correlator determinadas por la señal recibida $x(t)$ son los únicos datos útiles para el proceso de toma de decisiones y, por tanto, representan las *estadísticas suficientes* para el problema presente. Por definición, las estadísticas suficientes resumen la totalidad de la información importante proporcionada por un vector de observación.

Ahora podemos resumir los resultados presentados en esta sección, formulando el *teorema de irrelevancia:*

> En cuanto a lo que concierne a la detección de señales en ruido blanco gaussiano aditivo, sólo las proyecciones del ruido sobre las funciones básicas del conjunto de señales $\{s_i(t)\}_{i=1}^{M}$ afecta las estadísticas suficientes del problema de detección; el resto del ruido es irrelevante.

Como un corolario de este teorema, es posible establecer que el canal AWGN de la figura 5.2 es equivalente a un *canal de vectores N*-dimensional descrito por el vector de observación

$$\mathbf{x} = \mathbf{s}_i + \mathbf{w}, \quad i = 1, 2, \ldots, M \tag{5.48}$$

donde la dimensión N es el número de funciones básicas implicadas al formular el vector de señal $\mathbf{s}_i$. Las componentes individuales del vector de señal $\mathbf{s}_i$ y el vector de ruido $\mathbf{w}$ se definen mediante las ecuaciones (5.6) y (5.31), respectivamente. El teorema de irrelevancia y su corolario son realmente fundamentales para entender el problema de la detección de señales, como se describe a continuación.

5.4 *Funciones de verosimilitud*

Las funciones de densidad de probabilidad condicional $f_X(\mathbf{x}|m_i)$, $i = 1, 2, \ldots, M$, constituyen la verdadera caracterización del canal AWGN. Su deducción conduce a una dependencia funcional del vector de observación $\mathbf{x}$, produciendo el símbolo del mensaje transmitido m_i. Sin embargo, en el receptor tenemos una situación exactamente opuesta: se nos da el vector de observación $\mathbf{x}$ y se requiere estimar el símbolo del mensaje m_i responsable de generar $\mathbf{x}$. Para subrayar este último punto, introducimos la idea de una *función de verosimilitud* denotada por $L(m_i)$ y definida por

$$L(m_i) = f_X(\mathbf{x}|m_i), \qquad i = 1, 2, \ldots, M \tag{5.49}$$

Sin embargo, es importante admitir que a pesar de que $L(m_i)$ y $f_X(\mathbf{x}|m_i)$ tienen exactamente la misma forma matemática, sus significados individuales son diferentes.

En la práctica encontramos más conveniente trabajar con la *función de verosimilitud log* denotada por $l(m_i)$ y definida por

$$l(m_i) = \log L(m_i), \qquad i = 1, 2, \ldots, M \tag{5.50}$$

La función de verosimilitud log mantiene una relación uno a uno con la función de verosimilitud por dos razones:

1. Por definición, una función de densidad de probabilidad siempre es no negativa. En consecuencia, se concluye que la función de verosimilitud es similar a una cantidad no negativa.
2. La función logarítmica es una función monotónicamente creciente de su argumento.

El uso de la ecuación (5.46) en la (5.50) produce las funciones de verosimilitud log para un canal AWGN como

$$l(m_i) = -\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2, \qquad i = 1, 2, \ldots, M \tag{5.51}$$

donde hemos ignorado el término constante $-(N/2)\log(\pi N_0)$, puesto que no tiene relación alguna con el símbolo del mensaje m_i. Advierta que s_{ij}, $j = 1, 2, \ldots, N$, son los elementos del vector de señal $\mathbf{s}_i$ que representa al símbolo del mensaje m_i. Con la ecuación (5.51) a nuestra disposición, estamos listos para abordar el problema básico del diseño de receptores.

5.5 *Detección coherente de señales en ruido: decodificación de máxima verosimilitud*

Supóngase que en cada franja de tiempo de duración T segundos se transmite una de las M posibles señales $s_1(t), s_2(t), \ldots, s_M(t)$ con igual probabilidad, 1/M. En la representación geométrica de señales, la señal $s_i(t)$, $i = 1, 2, \ldots, M$, se aplica a un banco de correlatores, con una entrada común y alimentada por un conjunto apropiado de N funciones básicas ortonormales. Las salidas que se producen del correlator definen el *vector de señal* $\mathbf{s}_i$. Puesto que el conocimiento del *vector de señal* $\mathbf{s}_i$ es tan bueno como el de la propia señal transmitida $s_i(t)$, y viceversa, es posible representar $s_i(t)$ por un punto en un espacio euclidiano de dimensión $N \leq M$. Nos referimos a este punto como

el *punto de la señal transmitida* o *punto del mensaje*. El conjunto de puntos del mensaje correspondientes al conjunto de señales transmitidas $\{s_i(t)\}_{i=1}^{M}$ se denomina *constelación de la señal*.

Sin embargo, la representación de la señal recibida $x(t)$ se complica por la presencia de ruido aditivo $w(t)$. Advertimos que cuando la señal recibida $x(t)$ se aplica al banco de N correlatores, las salidas de los correlatores definen al vector de observación $\mathbf{x}$. De acuerdo con la ecuación (5.48), el vector $\mathbf{x}$ difiere del vector de señal $\mathbf{s}_i$ por el *vector de ruido* $\mathbf{w}$, cuya orientación es completamente aleatoria. El vector de ruido $\mathbf{w}$ está caracterizado por completo por el ruido $w(t)$; sin embargo, el inverso de este enunciado no es cierto. El vector de ruido $\mathbf{w}$ representa esa porción de ruido $w(t)$ que interferirá con el proceso de detección; la porción restante de este ruido, denotada por $w'(t)$, se sintoniza mediante el banco de correlatores.

Ahora bien, con base en el vector de observación $\mathbf{x}$, podemos representar la señal recibida $x(t)$ por un punto en el mismo espacio euclidiano que se usa para representar la señal transmitida. Nos referimos a este segundo punto como el *punto de la señal recibida*. La señal recibida vaga en torno al punto del mensaje de una manera totalmente aleatoria, porque puede ubicarse en cualquier punto dentro de una nube distribuida en forma gaussiana que se centra en el punto del mensaje. Lo anterior se ilustra en la figura 5.7*a* para el caso de un espacio de señal tridimensional. Para una realización particular del vector de ruido $\mathbf{w}$ (es decir, un punto particular dentro de la nube aleatoria de la figura 5.7*a*), la relación entre el vector de observación $\mathbf{x}$ y vector de señal $\mathbf{s}_i$ es como se ilustra en la figura 5.7*b*.

Estamos ahora listos para establecer el problema de la detección de señales:

> Dado el vector de observación $\mathbf{x}$, efectuar una correspondencia de $\mathbf{x}$ a una estimación $\hat{m}$ del símbolo transmitido m_i, de tal manera que minimice la probabilidad de error en el proceso de toma de decisiones.

Suponga que, dado el vector de observación $\mathbf{x}$, tomamos la decisión $\hat{m} = m_i$. La probabilidad de error en esta decisión, que denotamos mediante $P_e(m_i|\mathbf{x})$, es simplemente

$$\begin{aligned} P_e(m_i|\mathbf{x}) &= P(m_i \text{ no enviado}|\mathbf{x}) \\ &= 1 - P(m_i \text{ enviado}|\mathbf{x}) \end{aligned} \tag{5.52}$$

El criterio de toma de decisiones consiste en minimizar la probabilidad de error al hacer corresponder cada vector de observación $\mathbf{x}$ determinado con una decisión. Con base en la ecuación (5.52), es posible establecer la *regla de decisión óptima:*

$$\begin{gathered} \text{Sea } \hat{m} = m_i \text{ si} \\ P(m_i \text{ enviado}|\mathbf{x}) \geq P(m_k \text{ enviado}|\mathbf{x}) \quad \text{para toda } k \neq i \end{gathered} \tag{5.53}$$

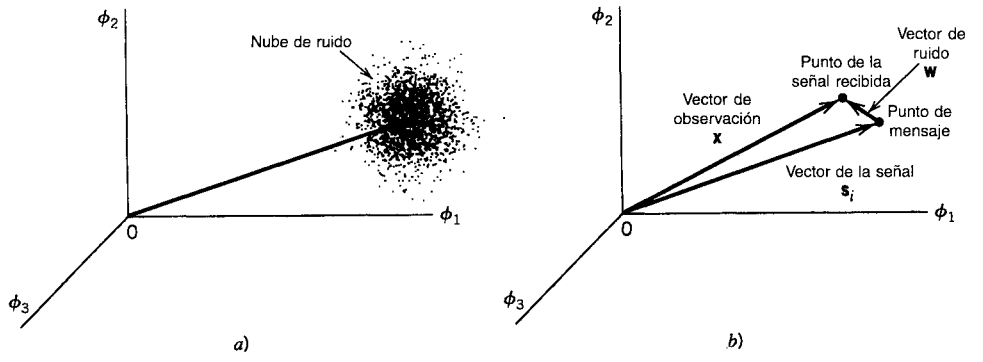


FIGURA 5.7 Ilustración del efecto de la perturbación del ruido, indicado en *a*, sobre la ubicación del punto de la señal recibida, indicado en *b*.

donde $k = 1, 2, \ldots, M$. Esta regla de decisión se conoce como la *regla de probabilidad máxima a posteriori (MAP)*.

Es posible expresar más explícitamente la condición de la ecuación (5.53) en términos de las probabilidades *a priori* de la señal transmitida y en términos de las funciones de probabilidad. Recurriendo a la regla de Bayes en la ecuación (5.53) e ignorando por ahora trabas en el proceso de toma de decisiones, podemos reenunciar la regla MAP del modo siguiente:

$$\begin{aligned} &\text{Sea } \hat{m} = m_i \text{ si} \\ &\frac{p_k f_{\mathbf{X}}(\mathbf{x}|m_k)}{f_{\mathbf{X}}(\mathbf{x})} \text{ es máximo para } k = i \end{aligned} \tag{5.54}$$

donde p_k es la probabilidad *a priori* de transmitir el símbolo m_k, $f_{\mathbf{X}}(\mathbf{x}|m_k)$ es la función de densidad de probabilidad condicional del vector de observación aleatorio $\mathbf{X}$ dada la transmisión del símbolo m_k, y $f_{\mathbf{X}}(\mathbf{x})$ es la función de densidad de probabilidad incondicional de $\mathbf{X}$. En la ecuación (5.54) es posible advertir lo siguiente:

- El término del denominador $f_{\mathbf{X}}(\mathbf{x})$ es independiente del símbolo transmitido.
- La probabilidad *a priori* $p_k = p_i$ cuando todos los símbolos de fuente se transmiten con igual probabilidad.
- La función de densidad de probabilidad condicional $f_{\mathbf{X}}(\mathbf{x}|m_k)$ mantiene una relación uno a uno con la función de verosimilitud logarítmica $l(m_k)$.

Por tanto, es posible reenunciar la regla de decisión de la ecuación (5.54) en términos de $l(m_k)$ simplemente del modo que sigue:

$$\begin{aligned} &\text{Sea } \hat{m} = m_i \text{ si} \\ &l(m_k) \text{ es máximo para } k = i \end{aligned} \tag{5.55}$$

Esta regla de decisión se conoce como la *regla de máxima verosimilitud*, y el dispositivo para su implementación se denomina en forma correspondiente *decodificador de máxima verosimilitud*. De acuerdo con la ecuación (5.55), un decodificador de máxima verosimilitud calcula las funciones de verosimilitud logarítmica como métricas para todos los M símbolos de mensajes posibles, los compara y luego decide en favor del máximo. De ese modo, el decodificador de máxima verosimilitud difiere del decodificador *a posteriori* máximo en que éste supone símbolos de mensajes igualmente probables.

Es útil contar con una interpretación gráfica de la regla de decisión de máxima verosimilitud. Dejemos que Z denote el espacio N-dimensional de todos los posibles vectores de observación $\mathbf{x}$. Nos referimos a este espacio como el *espacio de observación* debido a que hemos supuesto que la regla de decisión debe indicar $\hat{m} = m_i$, donde $i = 1, 2, \ldots, M$; el espacio de observación total Z se divide respectivamente en M *regiones de decisión*, denotadas por $Z_1, Z_2, \ldots, Z_M$. Por tanto, es posible reenunciar la regla de decisión de la ecuación (5.55) del modo siguiente:

$$\begin{aligned} &\text{El vector de observación } \mathbf{x} \text{ se encuentra en la región } Z_i \text{ si} \\ &l(m_k) \text{ es máximo para } k = i \end{aligned} \tag{5.56}$$

Además de las fronteras entre las regiones de decisión $Z_1, Z_2, \ldots, Z_M$, es claro que este conjunto de regiones cubre el espacio completo de posibles vectores de observación $\mathbf{x}$. Adoptamos la convención de que todas las trabas se resuelven al azar; esto es, el receptor simplemente adivina. Específicamente, si el vector de observación $\mathbf{x}$ cae sobre la frontera entre cualesquiera dos regiones de decisión, Z_i y Z_k, digamos, la elección entre las dos posibles decisiones $\hat{m} = m_i$ y $\hat{m} = m_k$ se resuelve *a priori* mediante el lanzamiento de una moneda sin alterar. Claramente, el resultado de un evento tal no afecta el valor final del error de probabilidad, puesto que en esta frontera la condición de la ecuación (5.53) se satisface con el signo de igualdad.

La regla de decisión de máxima verosimilitud de la ecuación (5.55) o su contraparte geométrica descrita en la ecuación (5.56) es de tipo genérico, siendo el ruido aditivo del canal $w(t)$ la única restricción que se le impone. A continuación especializamos esta regla para el caso en el que $w(t)$ es tanto blanco como gaussiano.

De la función de verosimilitud logarítmica definida en la ecuación (5.51) para un canal AWGN observamos que $l(m_k)$ alcanza su valor máximo cuando el término de la sumatoria

$$\sum_{j=1}^{N} (x_j - s_{kj})^2$$

se minimiza al elegir $k = i$. En consecuencia, es posible formular la regla de decisión de máxima verosimilitud para un canal AWGN como

El vector de observación $\mathbf{x}$ se encuentra en la región Z_i si

$$\sum_{j=1}^{N} (x_j - s_{kj})^2 \text{ es mínima para } k = i \tag{5.57}$$

Después de esto, advertimos con base en nuestra explicación anterior que [vea la ecuación (5.14) para comparar]

$$\sum_{j=1}^{N} (x_j - s_{kj})^2 = \| \mathbf{x} - \mathbf{s}_k \|^2 \tag{5.58}$$

donde $\|\mathbf{x} - \mathbf{s}_k\|$ es la distancia euclidiana entre el punto de la señal recibida y el punto del mensaje, representadas, respectivamente, por los vectores $\mathbf{x}$ y $\mathbf{s}_k$. Por ello, es posible reenunciar la regla de decisión de la ecuación (5.57) de la manera siguiente:

El vector de observación $\mathbf{x}$ se encuentra en la región Z_i si

la distancia euclidiana $\| \mathbf{x} - \mathbf{s}_k \|$ es mínima para $k = i$ (5.59)

La ecuación (5.59) establece que *la regla de decisión de máxima verosimilitud consiste simplemente en elegir el punto del mensaje lo más cerca del punto de la señal recibida*, lo que se satisface de modo intuitivo.

En la práctica, la necesidad de elevar al cuadrado la regla de decisión de la ecuación (5.59) se evita al admitir que

$$\sum_{j=1}^{N} (x_j - s_{kj})^2 = \sum_{j=1}^{N} x_j^2 - 2 \sum_{j=1}^{N} x_j s_{kj} + \sum_{j=1}^{N} s_{kj}^2 \tag{5.60}$$

El término de la primera sumatoria de esta expansión es independiente del índice k y, por tanto, es posible ignorarlo. El término de la segunda sumatoria es el producto interno del vector de observación $\mathbf{x}$ y el vector de señal $\mathbf{s}_k$. El término de la tercera sumatoria corresponde a la energía de la señal transmitida $s_k(t)$. En consecuencia, es posible formular una regla de decisión equivalente a la de la ecuación (5.59) de la siguiente forma:

El vector de observación $\mathbf{x}$ se encuentra en la región Z_i si

$$\sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2} E_k \text{ es máximo para } k = i \tag{5.61}$$

donde E_k es la energía de la señal transmitida $s_k(t)$:

$$E_k = \sum_{j=1}^{N} s_{kj}^2 \tag{5.62}$$

De acuerdo con la ecuación (5.61), deducimos que para un canal AWGN las regiones de decisión se encuentran en el espacio de observación N-dimensional Z, acotadas por fronteras lineales [$(N - 1)$-dimensional]. La figura 5.8 muestra el ejemplo de regiones de decisión para $M = 4$

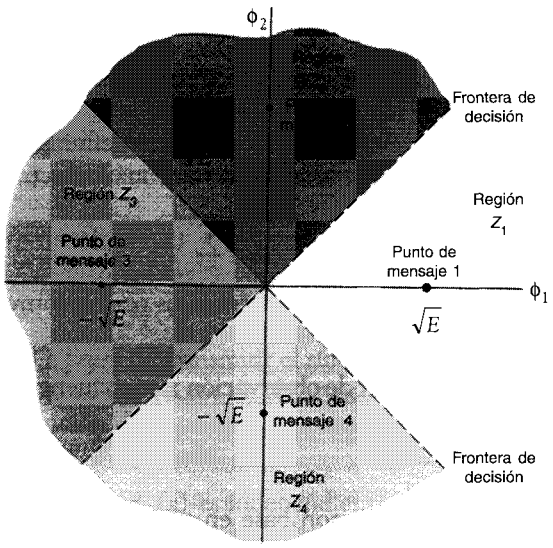


FIGURA 5.8 Ilustración de la división del espacio de observación en regiones de decisión en el caso en el que $N = 2$ y $M = 4$; se supone que los M símbolos transmitidos son igualmente probables.

señales y $N = 2$ dimensiones, suponiendo que las señales se transmiten con igual energía E, e igual probabilidad.

5.6 *Receptor de correlación*

Con base en el material que se presentó en las secciones anteriores, encontramos que para un canal AWGN y para el caso en el que las señales transmitidas $s_1(t), s_2(t), \ldots, s_M(t)$ son igualmente probables, el receptor óptimo está compuesto por dos subsistemas, mismos que se detallan en la figura 5.9 y que se describen a continuación:

1. La parte *detectora* del receptor se muestra en la figura 5.9*a*. Consiste en un banco de M *integradores de producto* o *correlatores*, suministrados con un conjunto correspondiente de señales de referencia coherentes o funciones básicas ortogonales $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$ que se generan localmente. Este banco de correlatores opera sobre la señal recibida $x(t)$, $0 \leq t \leq T$, para producir el vector de observación $\mathbf{x}$.
2. La segunda parte del receptor, a saber, el *decodificador de transmisión de la señal*, se muestra en la figura 5.9*b*. Éste se pone en práctica en la forma de un decodificador de máxima verosimilitud que opera sobre el vector de observación $\mathbf{x}$ para producir una estimación, $\hat{m}$, del símbolo transmitido m_i, $i = 1, 2, \ldots, M$ en una forma que minimizaría la probabilidad promedio del error de símbolo. De acuerdo con la ecuación (5.61), los N elementos del vector de observación $\mathbf{x}$ se multiplican primero por los correspondientes N elementos de cada uno de los M vectores de señal $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$, y los productos que se producen se suman sucesivamente en acumuladores para formar el conjunto correspondiente de productos interiores $\{\mathbf{x}^T\mathbf{s}_k | k = 1, 2, \ldots, M\}$. Después, los productos interiores se corrigen porque las energías de la señal transmitida quizá sean desiguales. Por último, se elige el más grande del conjunto resultante de números, y se toma una decisión apropiada acerca del mensaje transmitido.

El receptor óptimo de la figura 5.9 suele conocerse como un *receptor de correlación*.

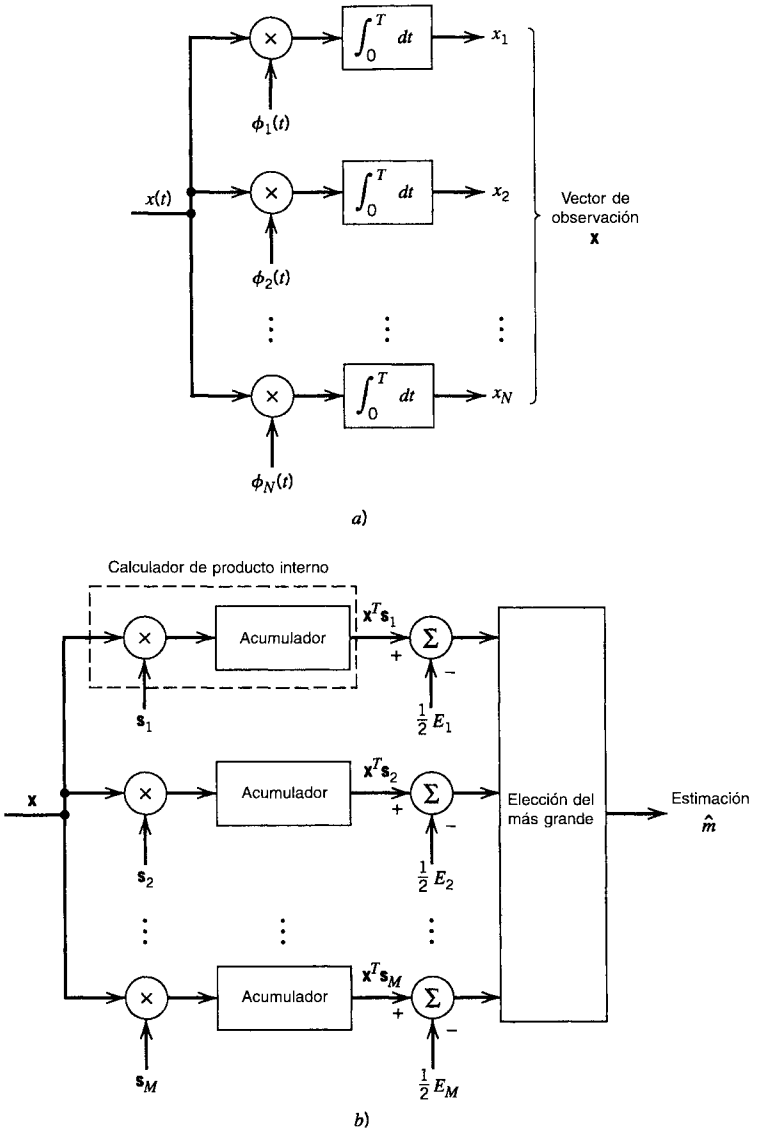


FIGURA 5.9 *a*) Detector o demodulador. *b*) Decodificador de transmisión de la señal.

■ EQUIVALENCIA DE RECEPTORES DE CORRELACIÓN Y DE FILTRO DE ACOPLAMIENTO

El receptor que se ilustra en la figura 5.9*a* incluye un conjunto de correlatores. De manera alternativa, podemos usar un conjunto correspondiente de *filtros de acoplamiento* para construir el detector; el filtro de acoplamiento y sus propiedades se consideraron en la sección 4.2. Para demostrar la equivalencia de un correlator de y un filtro de acoplamiento, considere un filtro lineal e invariante con el tiempo con respuesta al impulso $h_j(t)$. Con la señal recibida $x(t)$ utilizada como la entrada del filtro, la salida del filtro que se produce, $y_j(t)$, se define por medio de la integral de convolución:

$$y_j(t) = \int_{-\infty}^{\infty} x(\tau)h_j(t-\tau)d\tau \tag{5.63}$$

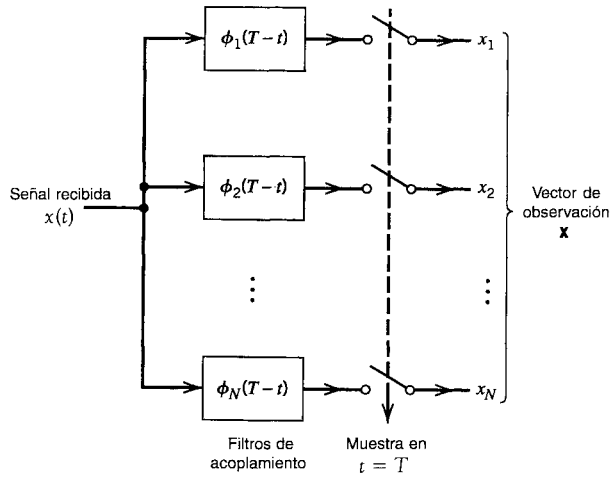


FIGURA 5.10 Parte del detector del receptor de filtro de acoplamiento; el decodificador de transmisión de la señal es como se muestra en la figura 5.9*b*.

De la definición de un filtro de acoplamiento que se presentó en la sección 4.2, recordamos que la respuesta al impulso $h_j(t)$ de un filtro lineal e invariante con el tiempo con una señal de entrada $\phi_j(t)$ es una versión invertida en el tiempo y retardada $\phi_j(t)$. Supóngase que establecemos

$$h_j(t) = \phi_j(T - t) \tag{5.64}$$

Entonces la salida del filtro resultante es

$$y_j(t) = \int_{-\infty}^{\infty} x(\tau)\phi_j(T - t + \tau)d\tau \tag{5.65}$$

Al muestrear esta salida en el tiempo $t = T$, obtenemos

$$y_j(T) = \int_{-\infty}^{\infty} x(\tau)\phi_j(\tau)d\tau$$

Puesto que, por definición $\phi_j(t)$ es cero fuera del intervalo $0 \le t \le T$, encontramos que $y_j(T)$ es en realidad la salida x_j del j-ésimo correlator producida por la señal recibida $x(t)$ en la figura 5.9*a*, de acuerdo con

$$y_j(T) = \int_0^T x(\tau)\phi_j(\tau)d\tau \tag{5.66}$$

Por tanto, la parte del detector del receptor óptimo tal vez también pueda ponerse en práctica utilizando un banco de filtros de acoplamiento, como se ilustra en la figura 5.10. Sin embargo, es importante advertir que la salida de cada correlator en la figura 5.9*a* es equivalente a la salida de un filtro de acoplamiento correspondiente en la figura 5.10 sólo cuando esa salida se muestrea en el tiempo $t = T$.

5.7 *Probabilidad de error*

Para completar la caracterización estadística del receptor de correlación que se describe en la figura 5.9, necesitamos evaluar su desempeño ante el ruido. Para hacerlo, suponemos que el espacio de observación Z se divide, de acuerdo con la regla de decisión de máxima verosimilitud, en un conjunto de M regiones $\{Z_i\}_{i=1}^{M}$. Suponga también que se transmite el símbolo m_i (o, de modo

equivalente, el vector de señal $\mathbf{s}_i$) y que se recibe un vector de observación $\mathbf{x}$. En ese caso ocurre un error siempre que el punto de la señal recibida representada por $\mathbf{x}$ no caiga dentro de la región Z_i asociada con el punto del mensaje representado por $\mathbf{s}_i$. Promediando sobre todos los posibles símbolos transmitidos, vemos fácilmente que la *probabilidad promedio del error del símbolo*, P_e, es

$$\begin{aligned} P_e &= \sum_{i=1}^{M} p_i \, P(\mathbf{x} \text{ no está enviado en } Z_i | m_i) \\ &= \frac{1}{M} \sum_{i=1}^{M} P(\mathbf{x} \text{ no está enviado en } Z_i | m_i) \\ &= 1 - \frac{1}{M} \sum_{i=1}^{M} P(\mathbf{x} \text{ está enviado en } Z_i | m_i) \end{aligned} \tag{5.67}$$

donde hemos utilizado la notación estándar para denotar la probabilidad y la probabilidad condicional de un evento. Puesto que $\mathbf{x}$ es el valor de muestreo de un vector aleatorio $\mathbf{X}$, es posible reescribir la ecuación (5.67) en términos de la función de verosimilitud (cuando se envía m_i) como sigue:

$$P_e = 1 - \frac{1}{M} \sum_{i=1}^{M} \int_{Z_i} f_{\mathbf{X}}(\mathbf{x}|m_i) \, d\mathbf{x} \tag{5.68}$$

Para un vector de observación N-dimensional, la integral en la ecuación (5.68) es de igual modo N-dimensional.

Invarianza de la probabilidad de error ante la rotación y la traslación

La manera en que se divide el espacio de observación Z en un conjunto de regiones $Z_1, Z_2, \ldots, Z_M$ en la detección de máxima verosimilitud de una señal en ruido blanco gaussiano aditivo, se define en forma armónica mediante la constelación de señal bajo estudio. En consecuencia, los cambios en la orientación de la constelación de la señal tanto de los ejes coordenados como del origen del espacio de señal *no* afecta la probabilidad del error de símbolo P_e definida en la ecuación (5.68). Este resultado es una consecuencia de dos hechos:

1. En la detección de máxima verosimilitud, la probabilidad de error del símbolo P_e depende únicamente de las distancias euclidianas relativas entre los puntos de mensaje en la constelación.
2. El ruido blanco gaussiano aditivo es *simétrico esféricamente* en todas las direcciones en el espacio de la señal.

Considere primero la invarianza de P_e con respecto a la rotación. El efecto de una rotación aplicada a todos los puntos de mensaje en una constelación es equivalente a multiplicar el vector de la señal N-dimensional $\mathbf{s}_i$ por una *matriz ortonormal* N por N denotada mediante $\mathbf{Q}$ para toda i. La matriz $\mathbf{Q}$ satisface la condición

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{I} \tag{5.69}$$

donde $\mathbf{I}$ es la *matriz de identidad* cuyos elementos diagonales son todos unitarios y sus elementos fuera de la diagonal son en su totalidad iguales a cero. Advierta que de acuerdo con la ecuación (5.69), la inversa de una matriz ortonormal de valores reales es igual a su forma transpuesta. De ese modo, el vector de señal $\mathbf{s}_i$ se sustituye por su versión rotada

$$\mathbf{s}_{i,\text{rotado}} = \mathbf{Q}\mathbf{s}_i, \quad i = 1, 2, \ldots, M \tag{5.70}$$

En forma correspondiente, el vector de ruido $\mathbf{w}$ de N por 1 se sustituye por su versión rotada

$$\mathbf{w}_{\text{rotado}} = \mathbf{Q}\mathbf{w} \tag{5.71}$$

Sin embargo, las características estadísticas del vector de ruido no son afectadas por esta rotación en virtud de las siguientes causas:

- Del capítulo 1 recordamos que una combinación lineal de variables aleatorias gaussianas también es gaussiana. Puesto que el vector de ruido $\mathbf{w}$ es gaussiano, por suposición, se sigue que el vector de ruido rotado $\mathbf{w}_{\text{rotado}}$ también es gaussiano.
- Puesto que el vector de ruido $\mathbf{w}$ tiene media cero, el vector de ruido $\mathbf{w}_{\text{rotado}}$ también tiene media cero, según indica

$$\begin{aligned} E[\mathbf{w}_{\text{rotado}}] &= E[\mathbf{Q}\mathbf{w}] \\ &= \mathbf{Q}E[\mathbf{w}] \\ &= \mathbf{0} \end{aligned} \tag{5.72}$$

- La matriz de covarianza de vector de ruido $\mathbf{w}$ es igual a $(N_0/2)\mathbf{I}$, donde $N_0/2$ es la densidad espectral de potencia del ruido blanco gaussiano aditivo $w(t)$; esto es,

$$E[\mathbf{w}\mathbf{w}^T] = \frac{N_0}{2}\mathbf{I} \tag{5.73}$$

Por tanto, la matriz de covarianza del vector de ruido rotado $\mathbf{w}_{\text{rotado}}$ es

$$\begin{aligned} E[\mathbf{w}_{\text{rotado}}\mathbf{w}_{\text{rotado}}^T] &= E[\mathbf{Q}\mathbf{w}(\mathbf{Q}\mathbf{w})^T] \\ &= E[\mathbf{Q}\mathbf{w}\mathbf{w}^T\mathbf{Q}^T] \\ &= \mathbf{Q}E[\mathbf{w}\mathbf{w}^T]\mathbf{Q}^T \\ &= \frac{N_0}{2}\mathbf{Q}\mathbf{Q}^T \\ &= \frac{N_0}{2}\mathbf{I} \end{aligned} \tag{5.74}$$

donde en las últimas dos líneas hemos utilizado las ecuaciones (5.73) y (5.69).

A la luz de estas anotaciones, es posible expresar el vector de observación para la constelación de señales rotada como

$$\mathbf{x}_{\text{rotado}} = \mathbf{Q}\mathbf{s}_i + \mathbf{w}, \qquad i = 1, 2, \ldots, M \tag{5.75}$$

De la ecuación (5.59) sabemos que la regla de decisión para la detección de máxima verosimilitud se basa en la distancia euclidiana del vector de observación $\mathbf{x}_{\text{rotado}}$ al vector de la señal rotada $\mathbf{s}_{i,\text{rotado}} = \mathbf{Q}\mathbf{s}_i$. Al comparar la ecuación (5.75) con la (5.48), vemos de inmediato que

$$\|\mathbf{x}_{\text{rotado}} - \mathbf{s}_{i,\text{rotado}}\| = \|\mathbf{x} - \mathbf{s}_i\| \qquad \text{para toda } i \tag{5.76}$$

Por tanto, podemos establecer formalmente el *principio de invarianza rotacional*:

> Si una constelación de señales se gira mediante una transformación ortonormal, esto es,
>
> $$\mathbf{s}_{i,\text{rotado}} = \mathbf{Q}\mathbf{s}_i, \quad i = 1, 2, \ldots, M$$
>
> donde $\mathbf{Q}$ es una matriz ortogonal, entonces queda por completo invariable la probabilidad del error de símbolo P_e en el que se incurre en la detección de señales de máxima verosimilitud sobre un canal AWGN.

Ilustramos este principio con un ejemplo. La constelación de señales que se muestra en la figura 5.11*b* es la misma que aparece en la figura 5.11*a*, excepto en que ésta se ha girado 45 grados. Aunque estas dos constelaciones en realidad se ven diferentes, el principio de invarianza rotacional nos indica de inmediato que la P_e es la misma para ambas.

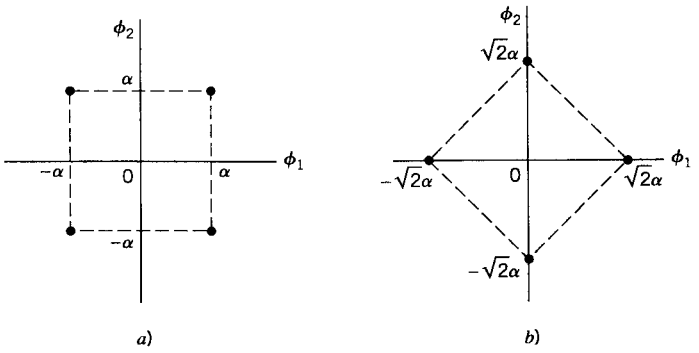


FIGURA 5.11 Un par de constelaciones de señales para ilustrar el principio de la invarianza rotacional.

Considere a continuación el aspecto de la invarianza con respecto a la traslación. Suponga que todos los puntos del mensaje en una constelación de señales se trasladan por medio de una cantidad vectorial constante **a**, de acuerdo con

$$\mathbf{s}_{i,\,\text{trasladado}} = \mathbf{s}_i - \mathbf{a}, \quad i = 1, 2, \ldots, M \tag{5.77}$$

El vector de observación se traslada en forma correspondiente en la misma cantidad vectorial, como indica

$$\mathbf{x}_{\text{trasladado}} = \mathbf{x} - \mathbf{a} \tag{5.78}$$

De las ecuaciones (5.77) y (5.78) vemos que el traslado **a** es común tanto al vector de señal trasladado $\mathbf{s}_i$ como al vector de observación trasladado **x**. En consecuencia, deducimos de inmediato que

$$\| \mathbf{x}_{\text{trasladado}} - \mathbf{s}_{i,\,\text{trasladado}} \| = \| \mathbf{x} - \mathbf{s}_i \| \qquad \text{para toda } i \tag{5.79}$$

y por consiguiente formulamos el *principio de invarianza traslacional* de la manera siguiente:

> Si una constelación de señales se traslada mediante una cantidad vectorial constante, entonces permanece completamente invariable la probabilidad del error de símbolo P_e en el que se incurre en la detección de señal de máxima verosimilitud sobre un canal AWGN.

Como ejemplo, considere las constelaciones de dos señales que se muestran en la figura 5.12, las cuales pertenecen a un par de señales PAM diferentes de cuatro niveles. La constelación de la figura 5.12*b* es la misma que aparece en la figura 5.12*a*, salvo por la traslación de $3\alpha/2$ hacia la derecha a lo largo del eje ϕ_1. El principio de invarianza traslacional afirma que el P_e es el mismo para ambas constelaciones.

SEÑALES DE ENERGÍA MÍNIMA

Una aplicación útil del principio de invarianza traslacional es la traslación de una constelación de señales determinadas de tal manera que se minimiza la energía promedio. Para investigar

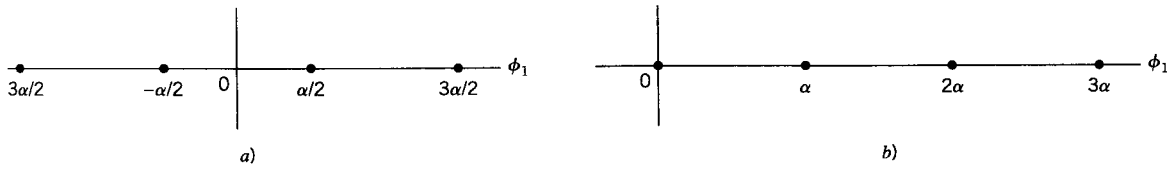


FIGURA 5.12 Un par de constelaciones de señales para ilustrar el principio de la invarianza traslacional.

este asunto, considere un conjunto de símbolos $m_1, m_2, \ldots, m_M$ representado por los vectores de señal $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$, respectivamente. La energía promedio de esta constelación de señales trasladada por una cantidad vectorial $\mathbf{a}$ es

$$\mathscr{E}_{\text{trasladado}} = \sum_{i=1}^{M} \|\mathbf{s}_i - \mathbf{a}\|^2 p_i \tag{5.80}$$

donde p_i es la probabilidad de que el símbolo m_i sea emitido por la fuente de información. La distancia euclidiana entre $\mathbf{s}_i$ y $\mathbf{a}$ se desarrolla como

$$\|\mathbf{s}_i - \mathbf{a}\|^2 = \|\mathbf{s}_i\|^2 - 2\mathbf{a}^T\mathbf{s}_i + \|\mathbf{a}\|^2$$

de ese modo, es posible reescribir la ecuación (5.80) en la forma expandida

$$\begin{aligned}\mathscr{E}_{\text{trasladado}} &= \sum_{i=1}^{M} \|\mathbf{s}_i\|^2 p_i - 2\sum_{i=1}^{M} \mathbf{a}^T\mathbf{s}_i p_i + \|\mathbf{a}\|^2 \sum_{i=1}^{M} p_i \\ &= \mathscr{E} - 2\mathbf{a}^T E[\mathbf{s}] + \|\mathbf{a}\|^2\end{aligned} \tag{5.81}$$

donde $\mathscr{E}$ es la energía promedio de la constelación de señales original y

$$E[\mathbf{s}] = \sum_{i=1}^{M} \mathbf{s}_i p_i \tag{5.82}$$

Al diferenciar la ecuación (5.81) con respecto al vector $\mathbf{a}$ y luego igualar a cero el resultado, encontramos rápidamente que la traslación minimizada es

$$\mathbf{a}_{\text{mín}} = E[\mathbf{s}] \tag{5.83}$$

La energía promedio mínima de la constelación de señales trasladadas de este modo es

$$\mathscr{E}_{\text{trasladado, mín}} = \mathscr{E} - \|\mathbf{a}_{\text{mín}}\|^2 \tag{5.84}$$

En estas condiciones es posible establecer el procedimiento para determinar la *traslación de energía mínima*:

> Dada una constelación de señales $\{s_i\}_{i=1}^{M}$, la constelación de señales correspondiente con energía promedio mínima se obtiene sustrayendo de cada vector de señal $\mathbf{s}_i$ en la constelación dada una cantidad igual al vector constante $E[\mathbf{s}]$, donde $E[\mathbf{s}]$ se define mediante la ecuación (5.82).

Recordando que la energía (o potencia) que se necesita para la transmisión de señales es un recurso fundamental, la traslación de energía mínima proporciona un método importante para trasladar una constelación de señales de interés de manera que se haga mínimo el requerimiento de energía. Por ejemplo, la constelación de la figura 5.12*a* tiene energía promedio mínima, en tanto que la de la figura 5.12*b* no.

■ COTA DE UNIÓN SOBRE LA PROBABILIDAD DE ERROR[2]

En canales AWGN, la formulación de la probabilidad promedio del error de símbolo, P_e, es directa en términos conceptuales. Simplemente escribimos P_e en forma integral sustituyendo la ecuación (5.46) en la (5.68). Sin embargo, el cálculo numérico de la integral resulta impráctico, excepto en unos cuantos casos simples (aunque importantes). Para superar esta dificultad computacional, es posible recurrir al uso de *cotas* que suelen ser adecuadas para predecir la relación señal a ruido (dentro de más o menos un decibel) que se requiere para mantener una tasa de error preestablecida. La aproximación a la integral que defina a P_e se realiza al simplificar la integral o la región de integración. Enseguida utilizaremos el último procedimiento para desarrollar una cota superior simple aunque útil que recibe el nombre de *cota de unión* como una

aproximación a la probabilidad promedio del error de símbolo para un conjunto de M señales (símbolos) igualmente probables en un canal AWGN.

Sea A_{ik}, con $(i, k) = 1, 2, \ldots, M$, el evento en el que el vector de observación $\mathbf{x}$ está más próximo al vector de señal $\mathbf{s}_k$ que a $\mathbf{s}_i$, cuando se envía el símbolo m_i (vector $\mathbf{s}_i$). La probabilidad condicional del error de símbolo cuando se envía el símbolo m_i, $P_e(m_i)$, es igual a la probabilidad de la *unión de los eventos*, $A_{i1}, A_{i2}, \ldots, A_{i,i-1}, A_{i,i-1}, \ldots, A_{i,M}$. De la teoría de probabilidades sabemos que *la probabilidad de una unión finita de eventos está sobreacotada por la suma de las probabilidades de los eventos constituyentes*. En consecuencia, es posible escribir

$$P_e(m_i) \leq \sum_{\substack{k=1 \\ k \neq i}}^{M} P(A_{ik}), \qquad i = 1, 2, \ldots, M \tag{5.85}$$

Esta relación se ilustra en la figura 5.13 para el caso de M = 4. En la figura 5.13*a* mostramos los cuatro puntos del mensaje y las regiones de decisión asociadas, donde se supone que el punto $\mathbf{s}_1$ representa al símbolo transmitido. En la figura 5.13*b* mostramos las tres descripciones constituyentes del espacio de la señal donde, en cada caso, se retienen el punto del mensaje transmitido $\mathbf{s}_1$ y otro punto de mensaje. De acuerdo con la figura 5.13*a*, la probabilidad del error de símbolo $P_e(m_i)$, es igual a la probabilidad de que el vector de observación $\mathbf{x}$ se encuentre en la región sombreada del diagrama bidimensional del espacio de la señal. Claramente, esta probabilidad es menor que la suma de las probabilidades de los tres eventos individuales en que $\mathbf{x}$ se encuentre en las regiones sombreadas de los tres espacios de señal constituyentes descritos en la figura 5.13*b*.

Es importante notar que, en general, la probabilidad $P(A_{ik})$ es diferente de la probabilidad $P(\hat{m} = m_k | m_i)$. La última es la probabilidad de que el vector de observación $\mathbf{x}$ se encuentre más

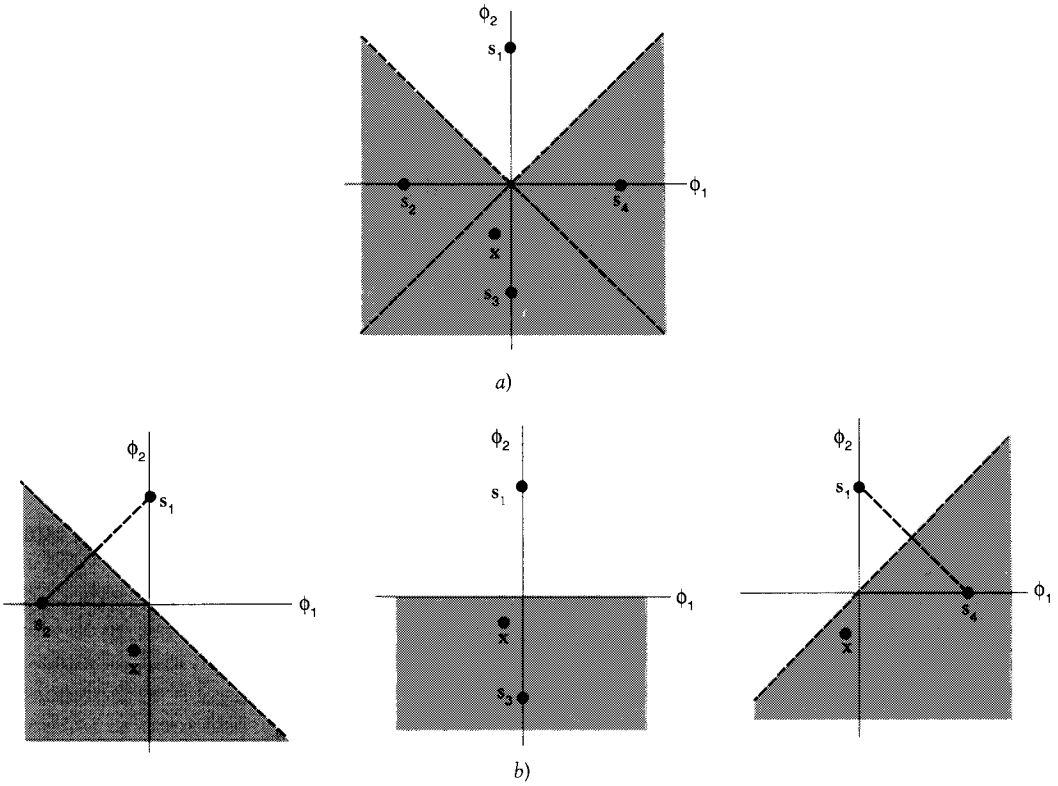


FIGURA 5.13 Ilustración de la cota de unión. *a*) Constelación de cuatro puntos de mensaje. *b*) Tres constelaciones con un punto de mensaje común y otro punto de mensaje retenido de la constelación original.

próximo al vector de señal $\mathbf{s}_k$ que cualquier otro, cuando se envía $\mathbf{s}_i$ (o m_i). Por otra parte, la probabilidad $P(A_{ik})$ depende sólo de dos vectores de señal, $\mathbf{s}_i$ y $\mathbf{s}_k$. Para subrayar esta diferencia, reescribimos la ecuación (5.85) adoptando $P_2(\mathbf{s}_i, \mathbf{s}_k)$ en lugar de $P(A_{ik})$. En consecuencia, escribimos

$$P_e(m_i) \leq \sum_{\substack{k=1 \\ k \neq i}}^{M} P_2(\mathbf{s}_i, \mathbf{s}_k), \qquad i = 1, 2..., M \tag{5.86}$$

La probabilidad $P_2(\mathbf{s}_i, \mathbf{s}_k)$ se denomina la *probabilidad de error de par dirigido* en el sentido de que si un sistema de transmisión de datos utiliza sólo un par de señales, $\mathbf{s}_i$ y $\mathbf{s}_k$, entonces $P_2(\mathbf{s}_i, \mathbf{s}_k)$ es la probabilidad de que el receptor confunda $\mathbf{s}_k$ por $\mathbf{s}_i$.

Considere entonces un sistema de comunicación simplificado que incluye el uso de dos mensajes igualmente probables representados por los vectores $\mathbf{s}_i$ y $\mathbf{s}_k$. Puesto que el ruido blanco gaussiano se distribuye de manera idéntica a lo largo de cualquier conjunto de ejes ortogonales, es posible elegir temporalmente el primer eje en un conjunto de este tipo como aquel que pasa por los puntos $\mathbf{s}_i$ y $\mathbf{s}_k$; en la figura 5.13*b* se presentan tres ejemplos. La frontera de decisión correspondiente se representa por medio del bisector que es perpendicular a la línea que une los puntos $\mathbf{s}_i$ y $\mathbf{s}_k$. Por tanto, cuando se envía el símbolo m_i (vector $\mathbf{s}_i$), y el vector de observación $\mathbf{x}$ se encuentra del lado del bisector donde yace $\mathbf{s}_k$, se produce un error. La probabilidad de este evento está dada por

$$P_2(\mathbf{s}_i, \mathbf{s}_k) = P(\mathbf{x} \text{ está mas próximo a } \mathbf{s}_k \text{ que a } \mathbf{s}_i, \text{cuando se envía } \mathbf{s}_i)$$

$$= \int_{d_{ik}/2}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{v^2}{N_0}\right) dv \tag{5.87}$$

donde d_{ik} es la distancia euclidiana entre $\mathbf{s}_i$ y $\mathbf{s}_k$; esto es,

$$d_{ik} = \|\mathbf{s}_i - \mathbf{s}_k\| \tag{5.88}$$

De la definición de la función de error complementaria tenemos

$$\text{ferc}(u) = \frac{2}{\sqrt{\pi}} \int_u^{\infty} \exp(-z^2)\, dz$$

Así, en términos de esta función, con z igual a $v/\sqrt{N_0}$, encontramos que la ecuación (5.87) adquiere la forma compacta

$$P_2(\mathbf{s}_i, \mathbf{s}_k) = \frac{1}{2} \text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \tag{5.89}$$

Al sustituir la ecuación (5.89) en la (5.86), obtenemos

$$P_e(m_i) \leq \frac{1}{2} \sum_{\substack{k=1 \\ k \neq i}}^{M} \text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right), \quad i = 1, 2, \ldots, M \tag{5.90}$$

La probabilidad del error de símbolo, promediada sobre todos los símbolos M, está sobreacotada del modo siguiente:

$$P_e = \sum_{i=1}^{M} p_i P_e(m_i)$$

$$\leq \frac{1}{2} \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \neq i}}^{M} p_i \, \text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \tag{5.91}$$

donde p_i es la probabilidad de transmitir el símbolo m_i.

Existen dos formas especiales de la ecuación (5.91) que encontraremos útiles en el capítulo 6 acerca de la transmisión de datos pasobanda:

1. Suponga que la constelación de señales es *simétrica circularmente* alrededor del origen. Entonces la probabilidad condicional del error $P_e(m_i)$ es la misma para toda i, en cuyo caso la ecuación (5.91) se reduce a

$$P_e \leq \frac{1}{2}\sum_{\substack{k=1 \\ k\neq i}}^{M} \text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \text{ para toda } i \tag{5.92}$$

2. Defina la distancia mínima de una constelación de señales, $d_{\text{mín}}$, como la distancia euclidiana más pequeña entre cualesquiera dos puntos de la señal transmitida en la constelación, según se indica mediante

$$d_{\text{mín}} = \min_{k\neq i} d_{ik} \quad \text{para toda } i \text{ y } k \tag{5.93}$$

Por tanto, al admitir que la función de error complementaria ferc(u) es monótonamente decreciente en su argumento u, puede escribirse

$$\text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \leq \text{ferc}\left(\frac{d_{\text{mín}}}{2\sqrt{N_0}}\right) \quad \text{para toda } i \text{ y } k \tag{5.94}$$

Por tanto, generalmente es posible simplificar la cota sobre la probabilidad promedio del error de símbolo en la ecuación (5.91) como

$$P_e \leq \frac{(M-1)}{2}\text{ferc}\left(\frac{d_{\text{mín}}}{2\sqrt{N_0}}\right) \tag{5.95}$$

La propia función de error complementaria está acotada como[3]

$$\text{ferc}\left(\frac{d_{\text{mín}}}{2\sqrt{N_0}}\right) \leq \frac{1}{\sqrt{\pi}}\exp\left(-\frac{d_{\text{mín}}^2}{4N_0}\right) \tag{5.96}$$

En consecuencia, podemos simplificar aún más la cota de unión sobre P_e dada en la ecuación (5.95) como

$$P_e \leq \frac{(M-1)}{2\sqrt{\pi}}\exp\left(-\frac{d_{\text{mín}}^2}{4N_0}\right) \tag{5.97}$$

La ecuación (5.97) muestra que para un canal AWGN preestablecido, la probabilidad promedio del error de símbolo P_e disminuye exponencialmente con la distancia mínima al cuadrado, $d_{\text{mín}}^2$.

■ Comparaciones de probabilidades de error de bit y de símbolo

Hasta ahora, la única figura de mérito que hemos utilizado para evaluar el desempeño ante el ruido de un sistema de transmisión pasobanda digital ha sido la probabilidad promedio del error de símbolo. Esta figura de mérito constituye la elección natural cuando se transmiten mensajes de longitud $m = \log_2 M$, tales como los símbolos alfanuméricos. Sin embargo, cuando el requerimiento es transmitir datos binarios, como los de computadora digital, muchas veces es más útil emplear otra figura de mérito conocida como la tasa de error de bits (BER). Aun cuando en general no existe relación única entre estas dos figuras de mérito, resulta afortunado que dicha relación pueda derivarse de dos casos de interés prácticos, según se explica a continuación.

Caso 1

En el primer caso suponemos que es posible efectuar la correspondencia entre símbolos binarios y M-arios de tal manera que las dos M-tuplas binarias correspondientes a cada par de símbolos adyacentes en el esquema de modulación M-ario difiera únicamente en una posición de bit. Esta restricción de correspondencia se satisface utilizando un *código Gray*. Cuando la probabilidad del error de símbolo P_e es aceptablemente pequeña, encontramos que la probabilidad de confundir un símbolo por cualquiera de los dos símbolos "más cercanos" es mucho mayor que cualquier otro tipo de error de símbolo. Asimismo, dado un error de símbolo, el número más probable de errores de bit es 1, sujeto a la restricción de correspondencia mencionada antes. Puesto que hay $\log_2 M$ bits por símbolo, se concluye que la probabilidad promedio del error de símbolo se relaciona con la tasa de error de bits de la manera siguiente:

$$\begin{aligned} P_e &= P\left(\bigcup_{i=1}^{\log_2 M} \{\text{el } i\text{ - ésimo bit está en error}\}\right) \\ &\leq \sum_{i=1}^{\log_2 M} P(\text{los } i\text{ - ésimo y } j\text{ - ésimo bits están en error}) \\ &= \log_2 M \cdot (\text{BER}) \end{aligned} \tag{5.98}$$

Advertimos también que

$$P_e \geq P(\text{el } i\text{ - ésimo bit está en error}) = \text{BER} \tag{5.99}$$

Por tanto, se concluye que la tasa de error de bits está acotada de la siguiente manera:

$$\frac{P_e}{\log_2 M} \leq \text{BER} \leq P_e \tag{5.100}$$

Caso 2

Sea $M = 2^K$, donde K es un entero. Suponemos que todos los errores de símbolo son igualmente probables y ocurren con probabilidad

$$\frac{P_e}{M-1} = \frac{P_e}{2^K - 1}$$

donde P_e es la probabilidad promedio del error de símbolo. ¿Cuál es la probabilidad de que el i-ésimo bit en un símbolo sea un error? Bien, hay 2^{K-1} casos de error de símbolo en los cuales se altera este bit particular, y existen 2^{K-1} casos en los que no se altera. Por tanto, la tasa de error de bits es

$$\text{BER} = \left(\frac{2^{K-1}}{2^K - 1}\right) P_e \tag{5.101}$$

o, de modo equivalente,

$$\text{BER} = \left(\frac{M/2}{M-1}\right) P_e \tag{5.102}$$

Advierta que para M grande, la tasa de error de bits se acerca al valor límite de $P_e/2$. La misma idea descrita aquí demuestra también que los errores de bit no son independientes, puesto que tenemos

$$P(\text{el i - ésimo bit está en error}) = \frac{2^{K-2}}{2^K - 1} P_e \neq (\text{BER})^2$$

5.8 *Resumen y discusión*

El objetivo fundamental del material que se presentó en este capítulo es la formulación de un procedimiento sistemático para el análisis y el diseño de un receptor de comunicación digital en la presencia de *ruido blanco gaussiano aditivo* (AWGN). El procedimiento, conocido como *detección de máxima verosimilitud*, decide cuál símbolo transmitido particular es el más probable de ocasionar la señal ruidosa que se observa en el canal de salida. El método que conduce a la formulación del detector de máxima verosimilitud (receptor) se denomina *análisis espacial de señales*. La idea básica del método es representar cada miembro de un conjunto de señales transmitidas mediante un vector N-dimensional donde N es el número de funciones básicas ortonormal necesarias para una representación geométrica única de las señales transmitidas. El conjunto de vectores de señal formado de ese modo define una *constelación de señales* en un espacio de señales N-dimensional.

Para una constelación de señales dada, la probabilidad (promedio) del error de símbolo P_e en el que se incurre en la detección de la señal de máxima verosimilitud sobre un canal AWGN es invariante ante la rotación de la constelación de señales, así como ante su traslación. Sin embargo, excepto por unos cuantos casos simples (pero importantes), el cálculo numérico de P_e es una propuesta impráctica. Para superar esta dificultad, la práctica usual consiste en recurrir al uso de cotas que llevan por sí solas a un cálculo directo. En este contexto, describimos la *cota de unión* que surge directamente del diagrama señal-espacio. La cota de unión se basa en una idea que se satisface intuitivamente: la probabilidad del error de símbolo P_e está dominada por los vecinos cercanos a la señal transmitida. Los resultados que se obtienen utilizando la cota de unión suelen ser bastante exactos cuando es elevada la relación señal a ruido.

Al disponer del material acerca del análisis espacial de señales y de los aspectos relacionados, estamos bien equipados para estudiar los sistemas de transmisión pasobanda, lo cual llevaremos a cabo en el capítulo 6.

NOTAS Y BIBLIOGRAFÍA

1. La representación geométrica de señales fue desarrollada primero por Kotel'nikov en 1947: V.A. Kotel'nikov, *The Theory of Optimum Noise Immunity* (Dover Publications), es una traducción de la disertación doctoral original presentada en enero de 1947 ante el consejo académico del Instituto de Energía Molotov en Moscú. En particular, vea la parte II del libro. Este método fructificó después por completo en el libro clásico de Wozencraft y Jacobs [172]. El análisis espacial de señales se analiza también en Cioffi [30], Anderson [7] y Proakis [123].

2. En la sección 5.7 obtuvimos la cota de unión para la probabilidad promedio del error de símbolo; la referencia clásica para esta cota es la de Wozencraft y Jacobs [172]. Para la derivación de cotas más estrictas vea Viterbi y Omura ([162], pp. 58-59).

3. En el capítulo 4 usamos la siguiente cota superior sobre la función de error complementaria

$$\operatorname{ferc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi u}}$$

Para u grande y positiva se obtiene una segunda cota para la función de error complementaria omitiendo el factor multiplicador $1/u$ en la cota superior anterior, de acuerdo con

$$\operatorname{ferc}(u) < \frac{\exp(-u^2)}{\sqrt{\pi}}$$

Esta segunda cota es superior a la que se usa en la ecuación (5.97).

PROBLEMAS

Representación de señales

5.1 En la sección 3.7 describimos códigos de línea para la modulación por codificación de pulsos. Con referencia al material que se presentó ahí, formule las constelaciones de señales para los siguientes códigos de línea:

a) Código unipolar sin retorno a cero
b) Código polar sin retorno a cero
c) Código unipolar con retorno a cero
d) Código Manchester

5.2 Una señal PAM de ocho niveles está definida por

$$s_i(t) = A_i \text{rect}\left(\frac{t}{T} - \frac{1}{2}\right)$$

donde $A_i = \pm 1, \pm 3, \pm 5, \pm 7$. Formule la constelación de señales de $\{s_i(t)\}_{i=1}^{8}$.

5.3 la figura P5.3 presenta las formas de onda de cuatro señales, $s_1(t)$, $s_2(t)$, $s_3(t)$ y $s_4(t)$.

a) Empleando el procedimiento de ortogonalización de Gram-Schmidt, encuentre una base ortonormal para este conjunto de señales.
b) Construya el diagrama correspondiente señal-espacio.

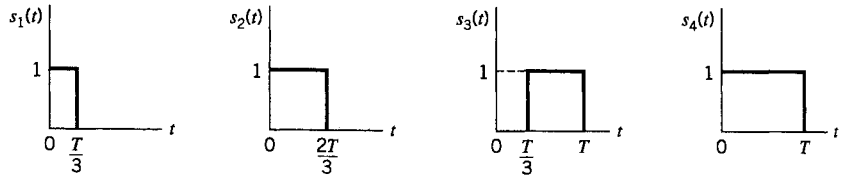


FIGURA P5.3

5.4 **a)** Utilizando el procedimiento de Gram-Schmidt, determine un conjunto de funciones básicas ortonormal para representar las tres señales $s_1(t)$, $s_2(t)$ y $s_3(t)$ que se muestran en la figura P5.4.

b) Exprese cada una de estas señales en términos del conjunto de funciones básicas que se encontraron en la parte *a*.

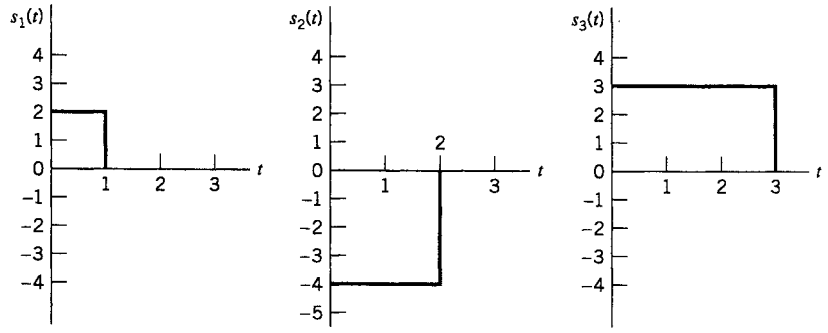


FIGURA P5.4

5.5 Un *conjunto de señales ortogonales* está caracterizado por la propiedad de que el producto interno de cualquier par de señales en el conjunto es igual a cero. La figura P5.5 muestra un par de señales $s_1(t)$ y $s_2(t)$ que cumplen esta condición. Construya la constelación de señales para $s_1(t)$ y $s_2(t)$.

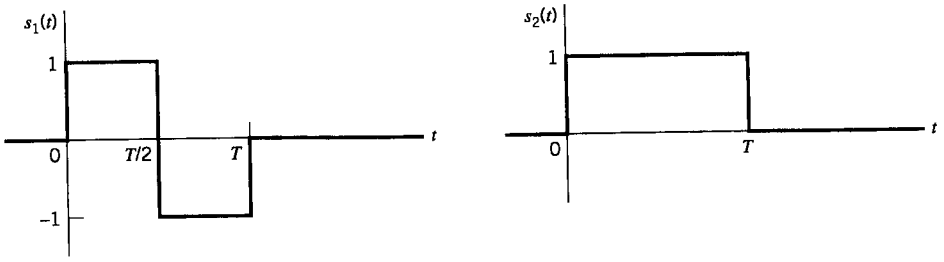


FIGURA P5.5

5.6 Una fuente de información emite un conjunto de símbolos denotado por $\{m_i\}_{i=1}^{M}$. Se consideran como candidatos dos esquemas de modulación, a saber, la modulación por duración de pulso (PDM) y la modulación por posición de pulso (PPM), para la representación eléctrica de este conjunto de símbolos. En la PDM, el i-ésimo símbolo se representa por medio de un símbolo de amplitud unitaria y duración $(i/M)T$. Por otra parte, en la PPM el i-ésimo símbolo se representa mediante un pulso corto de amplitud unitaria y duración fija, que se transmite en el tiempo $t = (i/M)T$. Demuestre que la PPM es la única de las dos que puede producir un conjunto ortogonal de señales sobre el intervalo $0 \le t \le T$.

5.7 Un conjunto de $2M$ *señales biortogonales* se obtiene de un conjunto de M señales ortogonales al aumentarlo con la negativa de cada señal en el conjunto.

a) La extensión de señales ortogonales a biortogonales deja sin cambio la dimensionalidad del espacio de señales. ¿Por qué?

b) Construya la constelación de señales correspondiente a las señales biortogonales del par de señales ortogonales que se muestran en la figura P5.5.

5.8 a) Un par de señales $s_i(t)$ y $s_k(t)$ tienen una duración común T. Demuestre que el producto interno de este par de señales está dado por

$$\int_0^T s_i(t)s_k(t)dt = \mathbf{s}_i^T \mathbf{s}_k$$

donde $\mathbf{s}_i$ y $\mathbf{s}_k$ son las representaciones vectoriales de $s_i(t)$ y $s_k(t)$, respectivamente.

b) Como continuación de la parte *a*, demuestre que

$$\int_0^T (s_i(t) - s_k(t))^2 dt = \|\mathbf{s}_i - \mathbf{s}_k\|^2$$

5.9 Considere un par de señales de valor complejo $s_1(t)$ y $s_2(t)$ que se representan respectivamente por

$$s_1(t) = a_{11}\phi_1(t) + a_{12}\phi_2(t), \quad -\infty < t < \infty$$
$$s_2(t) = a_{21}\phi_1(t) + a_{22}\phi_2(t), \quad -\infty < t < \infty$$

donde las funciones básicas $\phi_1(t)$ y $\phi_2(t)$ son ambas de valor real, aunque los coeficientes a_{11}, a_{12}, a_{21} y a_{22} son de valor complejo. Demuestre la forma compleja de la desigualdad de Schwarz:

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)dt\right|^2 \le \int_{-\infty}^{\infty} |s_1(t)|^2 dt \int_{-\infty}^{\infty} |s_2(t)|^2 dt$$

donde el asterisco denota conjugación compleja. ¿Cuándo se satisface esta relación con el signo de igualdad?

Proceso aleatorio

5.10 Considere un proceso aleatorio $X(t)$ que se expande en la forma

$$X(t) = \sum_{j=1}^{N} X_j \phi_j(t) + W'(t), \qquad 0 \le t \le T$$

donde $W'(t)$ es el término del ruido remanente. El $\{\phi_j(t)\}_{j=1}^N$ forma un conjunto ortonormal sobre el intervalo $0 \le t \le T$, y las X_j se definen mediante

$$X_j = \int_0^T X(t)\phi_j(t)dt$$

Sea $W'(t_k)$ una variable aleatoria que se obtiene al observar $W'(t)$ en el tiempo $t = t_k$. Demuestre que

$$E[X_j W'(t_k)] = 0, \quad \begin{cases} j = 1, 2, \ldots, N \\ 0 \le t_k \le T \end{cases}$$

5.11 Considere la detección óptima de la señal senoidal

$$s(t) = \text{sen}\left(\frac{8\pi t}{T}\right), \quad 0 \le t \le T$$

en ruido blanco gaussiano aditivo.

a) Determine la salida de correlator suponiendo una entrada sin ruido.

b) Determine la salida del correspondiente filtro de acoplamiento, considerando que éste incluye un retraso T para que sea causal.

c) Por tanto, demuestre que las dos salidas son las mismas sólo en el instante $t = T$.

Probabilidad de error

5.12 La figura P5.12 muestra un par de señales $s_1(t)$ y $s_2(t)$ que son ortogonales entre sí sobre el intervalo de observación $0 \le t \le 3T$. La señal recibida está definida por

$$x(t) = s_k(t) + w(t), \quad \begin{matrix} 0 \le t \le 3T \\ k = 1, 2 \end{matrix}$$

donde $w(t)$ es el ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$.

a) Diseñe un receptor que decida a favor de señales $s_1(t)$ o $s_2(t)$, suponiendo que estas dos señales son igualmente probables.

b) Calcule la probabilidad promedio del error de símbolo en el que se incurre por medio de este receptor para $E/N_0 = 4$, donde E es la energía de la señal.

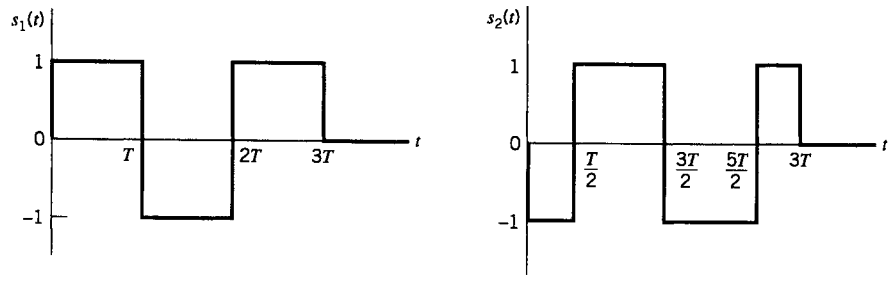


FIGURA P5.12

5.13 En el código Manchester, el símbolo binario 1 se representa mediante el pulso doble $s(t)$ que se muestra en la figura P5.13, y el símbolo binario 0 mediante el negativo de este pulso. Obtenga la fórmula para la probabilidad de error en el que se incurre mediante el procedimiento de detección de máxima verosimilitud aplicado a esta forma de transmisión de señales sobre un canal AWGN.

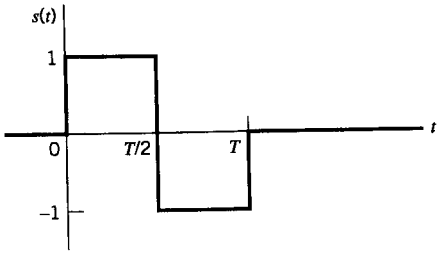


FIGURA P5.13

5.14 En la *prueba de Bayes*, aplicada a un problema de prueba de hipótesis primaria donde tenemos que elegir una de dos posibles hipótesis, H_0 y H_1, minimizamos el *riesgo*, definido por R

$$\begin{aligned} R = {} & C_{00}p_0P(\text{se dice } H_0|H_0 \text{ es verdadera}) \\ & + C_{10}p_0P(\text{se dice } H_1|H_0 \text{ es verdadera}) \\ & + C_{11}p_1P(\text{se dice } H_1|H_1 \text{ es verdadera}) \\ & + C_{01}p_1P(\text{se dice } H_0|H_1 \text{ es verdadera}) \end{aligned}$$

Los términos C_{00}, C_{10}, C_{11} y C_{01} denotan el costo asignado a los cuatro posibles resultados del experimento: el primer subíndice indica la hipótesis elegida y el segundo la hipótesis que es verdadera. Suponga que $C_{10} > C_{00}$ y $C_{01} > C_{11}$. Los términos p_0 y p_1 denotan, respectivamente, probabilidades *a priori* de las hipótesis H_0 y H_1.

a) Dado el vector de observación **x,** demuestre que la división del espacio de observación, a manera que se minimice el riesgo R, conduce a la *prueba de la proporción de verosimilitud*:

$$\begin{aligned} &\text{se dice } H_0 \text{ si } \Lambda(\mathbf{x}) < \lambda \\ &\text{se dice } H_1 \text{ si } \Lambda(\mathbf{x}) > \lambda \end{aligned}$$

donde $\Lambda(\mathbf{x})$ es la *proporción de verosimilitud.*

$$\Lambda(\mathbf{x}) = \frac{f_X(\mathbf{x}|H_1)}{f_X(\mathbf{x}|H_0)}$$

y λ es el *umbral* de la prueba definido por

$$\lambda = \frac{p_0(C_{10} - C_{00})}{p_1(C_{01} - C_{11})}$$

b) ¿Cuáles son los valores de costo para los cuales el criterio de Bayes se reduce a la probabilidad mínima del criterio del error?

Principios de invarianza rotacional y traslacional

5.15 Continuando con los cuatro códigos de línea considerados en el problema 5.1, identifique los que tienen la energía promedio mínima y los que no. Compare sus respuestas con la observaciones hechas sobre estos códigos de línea en la sección 3.7.

5.16 Considere las dos constelaciones que se muestran en la figura 5.11. Determine la matriz $\mathbf{Q}$ ortonormal que transforma la constelación indicada en la figura 5.11*a* en la que se presenta en la figura 5.11*b*.

5.17 a) La constelación de dos señales que se muestra en la figura 5.17 presenta la misma probabilidad promedio del error de símbolo. Justifique la validez de esta afirmación.

b) ¿Cuál de estas dos constelaciones tiene una energía promedio mínima? Justifique su respuesta. Puede suponer que los símbolos pertenecientes a los puntos de mensaje que se presentan en la figura P5.17 son igualmente probables.

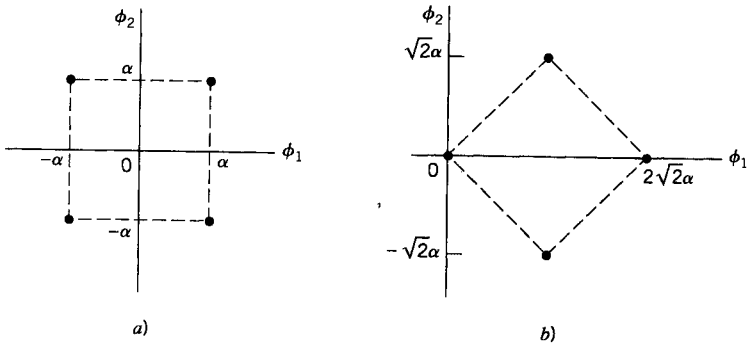


FIGURA P5.17

5.18 *Las señales simples (transortogonales)* son señales altamente correlacionadas y de igual probabilidad con la correlación más negativa que puede lograrse con un conjunto de M señales ortogonales. Esto es, el coeficiente de correlación entre cualquier par de señales en el conjunto se define por medio de

$$\rho_{ij} = \begin{cases} 1 & \text{para } i = j \\ \dfrac{-1}{M-1} & \text{para } i \neq j \end{cases}$$

Un método para la construcción de señales simples consiste en empezar con un conjunto de M señales ortogonales, cada una con energía E, y en aplicar después la traslación de energía mínima.

Considere un conjunto de tres símbolos igualmente probables cuya constelación de señales está compuesta por los vértices de un triángulo equilátero. Demuestre que estos tres símbolos integran un código simple.

Cotas en el error de probabilidad

5.19 En este problema investigamos las aproximaciones a la probabilidad de un error, P_e, para el par de señales opuestas que se muestra en la figura P5.19 ante la presencia de ruido blanco gaussiano aditivo de densidad espectral de potencia $N_0/2$. La fórmula exacta para P_e es

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right)$$

(Esta fórmula se deduce en la sección 6.3)

a) Utilizando las dos cotas superiores para la función de error complementaria dada en la nota 3, obtenga las aproximaciones correspondientes a P_e.

b) Compare las aproximaciones que se obtuvieron en la parte *a* para P_e con la fórmula exacta para $E_b/N_0 = 9$. Para el cálculo exacto de P_e, puede ampliar la tabla A6.6 acerca de la función de error.

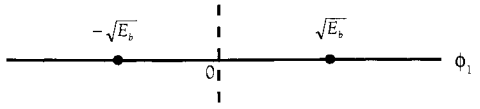


FIGURA P5.19

5.20 Considere el caso especial de una constelación de señales que tiene una *geometría simétrica* con respecto al origen. Suponga que los M puntos de mensaje de la constelación, pertenecientes a los símbolos $m_1, m_2, \ldots, m_M$, son igualmente probables. Utilizando la cota superior sobre la función de error complementaria dada en la ecuación (5.94), muestre que la probabilidad promedio del error de símbolo para la constelación está acotada como

$$P_e \leq \frac{M_{\text{mín}}}{2\sqrt{\pi}} \exp\left(-\min_{\substack{i,k \\ i \neq k}}\left(\frac{d_{ik}^2}{4N_0}\right)\right)$$

donde d_{ik} es la distancia euclidiana entre los puntos de mensaje i y k, y $M_{\text{mín}}$ es el número de señales transmitidas que alcanzan la mínima distancia euclidiana para cada m_i.

CAPÍTULO 6

TRANSMISIÓN DE DATOS PASOBANDA

Este capítulo se fundamenta en el material que se presentó en el capítulo 5 acerca del análisis del espacio de las señales. Expone el tema de la transmisión de datos digitales por el canal pasobanda que puede ser lineal o no lineal. Como en el caso de las comunicaciones analógicas, este modo de transmisión de datos se apoya en el uso de una onda portadora senoidal modulada por la secuencia de datos.

Específicamente, se abordan los siguientes temas:

- *Diferentes métodos de modulación digital, a saber, corrimiento de fase por llaveo, modulación en cuadratura y corrimiento de frecuencia por llaveo, así como sus variantes individuales.*
- *Detección coherente de señales moduladas en ruido blanco gaussiano aditivo, lo cual requiere que el receptor esté sincronizado con el transmisor en relación tanto con la fase portadora como con la temporización de bits.*
- *La detección no coherente de señales moduladas en ruido blanco gaussiano aditivo, descartando la información de fase en la señal recibida.*
- *Módems para la transmisión y recepción de datos digitales por la red pública de conmutación telefónica.*
- *Técnicas de modulación complejas, a saber, modulación de amplitud/fase sin portadora y multitonos discretos, para la transmisión de datos por un canal de banda ancha con medios para la severa interferencia de intersímbolo.*
- *Técnicas para sincronizar el receptor con el transmisor.*

6.1 *Introducción*

En la *transmisión por pulsos de banda base* que se estudió en el capítulo 4, una secuencia de datos representada en la forma de una señal discreta modulada por amplitud de pulsos (PAM) se transmite directamente por un canal pasobajas. En cambio, en la *transmisión pasobanda digital*, la secuencia entrante de datos se modula sobre una portadora (usualmente senoidal) con límites de frecuencia fijos impuestos por el canal pasobanda de interés; la transmisión de datos pasobanda se estudia en este capítulo.

El canal de comunicación que se utiliza para la transmisión de datos pasobanda puede ser un enlace de radio de microondas, un canal de satélite o similares. Incluso otras aplicaciones de la transmisión de datos pasobanda se encuentran en el diseño de códigos de línea pasobanda que se utilizan en lazos digitales de suscriptor y en técnicas de multiplexado ortogonal por división de frecuencia para la transmisión. En cualquier caso, el proceso de modulación que posibilita la transmisión implica la conmutación (llaveo) de la amplitud, la frecuencia o la fase de una portadora senoidal de acuerdo con los datos entrantes. Por tanto, existen tres esquemas básicos de transmisión de señales, los cuales se conocen como *corrimiento de amplitud por llaveo* (ASK), *corrimiento de frecuencia por llaveo* (FSK) y *corrimiento de fase por llaveo* (PSK). Éstos

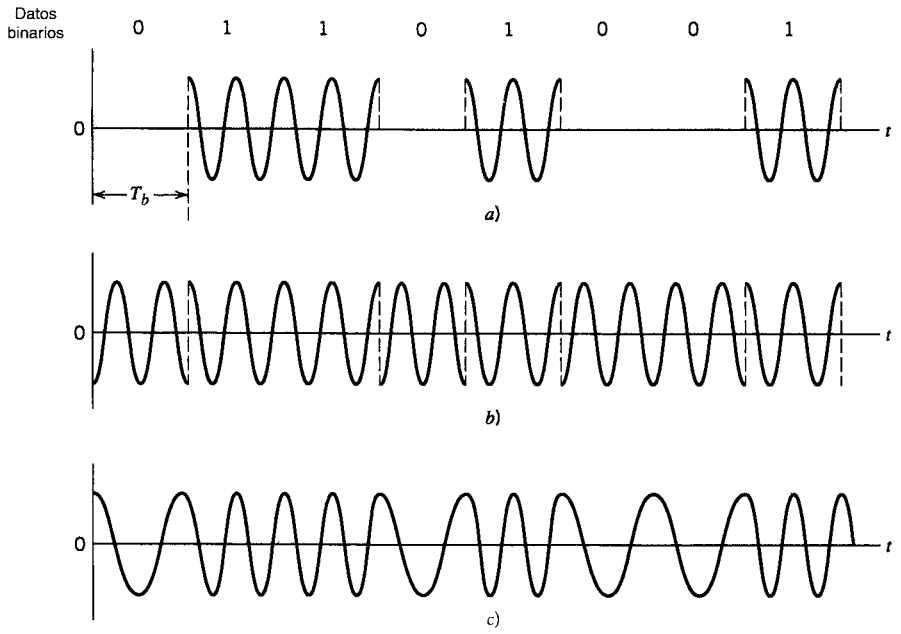


FIGURA 6.1 Formas de onda ilustrativas para las tres formas básicas de información binaria para transmisión de señales. *a*) Corrimiento de amplitud por llaveo. *b*) Corrimiento de fase por llaveo. *c*) Corrimiento de frecuencia por llaveo con fase continua.

pueden considerarse casos especiales en la modulación de amplitud, la modulación de frecuencia y la modulación de fase, respectivamente.

La figura 6.1 ilustra estos tres métodos de modulación para el caso de una fuente que suministra datos binarios. Vale la pena hacer notar los siguientes puntos de la figura 6.1:

- Aun cuando en la modulación de onda continua suele ser difícil distinguir entre las señales moduladas en fase y las moduladas en frecuencia al observar sus formas de onda, esto no es cierto para las señales PSK y FSK.
- A diferencia de las señales ASK, tanto las PSK como las FSK tienen una envolvente constante.

Esta última propiedad hace que las señales PSK y FSK sean impermeables a las no linealidades de amplitud, que suelen encontrarse en los canales de radio de microondas y de satélite. Es por esta razón que en la práctica encontramos que se prefieren las señales PSK y FSK sobre las ASK para la transmisión de datos pasobanda por canales no lineales.

JERARQUÍA DE LAS TÉCNICAS DE MODULACIÓN DIGITAL[1]

Las técnicas de modulación digital pueden clasificarse en *coherentes* y *no coherentes*, dependiendo de si el receptor está o no equipado con un circuito de recuperación de fase. El circuito de recuperación de fase asegura que el oscilador que suministra la onda portadora generada localmente en el receptor esté sincronizado (tanto en frecuencia como en fase) con el oscilador que provee la onda portadora que se emplea para modular originalmente la secuencia de datos entrante en el transmisor.

Como se explicó en el capítulo 4, en un *esquema de transmisión de señales M-aria* es posible enviar una de las cualesquiera M posibles señales $s_1(t), s_2(t), \ldots, s_M(t)$, durante cada intervalo de transmisión de señal de duración T. En casi todas las aplicaciones, el número de señales posibles

es $M = 2^n$, donde n es un entero. La duración del símbolo es $T = nT_b$, donde T_b es la duración de bit. En la transmisión de datos pasobanda, estas señales se generan cambiando la amplitud, fase o frecuencia de una portadora senoidal en M escalones discretos. De ese modo tenemos esquemas de modulación digital *ASK M-aria*, *PCK M-aria* y *FSK M-aria*. Otra manera de generar señales M-arias consiste en combinar diferentes métodos de modulación en una forma híbrida. Por ejemplo, podemos combinar cambios discretos tanto en la amplitud como en la fase portadora para producir la *amplitud-fase por llaveo M-aria* (APK). Una forma especial de esta modulación híbrida es la *modulación en amplitud por cuadratura* (QAM) M-aria, la cual tiene algunas propiedades atractivas. La ASK M-aria es un caso especial de la QAM M-aria.

Se prefieren los esquemas de transmisión de señales M-arias sobre los esquemas binarios para transmitir información digital por canales pasobanda cuando el requerimiento es conservar el ancho de banda a costa del incremento de potencia. En la práctica rara vez encontramos un canal de comunicación que tiene el ancho de banda exacto requerido para transmitir la salida de una fuente de información por medio de esquemas de transmisión de señales binarias. De ese modo, cuando el ancho de banda del canal es menor que el valor requerido, utilizamos esquemas de transmisión de señales M-arias para la máxima eficiencia. Con el fin de ilustrar la capacidad de conservación del ancho de banda de los esquemas de transmisiones M-arias, consideramos la transmisión de información consistente en una secuencia binaria con duración de bit T_b. Si vamos a transmitir esta información por medio de la PSK binaria, por ejemplo, requeriríamos un ancho de banda inversamente proporcional a T_b. Sin embargo, si tomamos bloques de n bits y usamos un esquema PSK M-ario, con $M = 2^n$ y duración de símbolo $T = nT_b$, el ancho de banda que se requiere es proporcional a $1/nT_b$. Esto demuestra que el uso de señales PSK M-arias permite una reducción en el ancho de banda de transmisión por un factor $n = \log_2$ sobre la PSK binaria.

La PSK y la QAM M-arias son ejemplos de *modulación lineal*. Sin embargo, difieren una de otra en un aspecto importante: una señal PSK M-aria tiene una envolvente constante, en tanto que una señal QAM M-aria cambia en la amplitud de la portadora. Por tanto, la PSK M-aria puede utilizarse para transmitir datos digitales por un canal pasobanda no lineal, en tanto que la QAM M-aria requiere el uso de un canal lineal.

La PSK M-aria, la QAM M-aria y la FSK M-aria suelen usarse en sistemas coherentes. El corrimiento de amplitud por llaveo y el corrimiento de frecuencia por llaveo conducen por sí solos en forma natural al uso de sistemas no coherentes, siempre que sea práctico mantener la sincronización de la fase portadora. Sin embargo, en el caso del corrimiento de fase por llaveo, no es posible tener "PSK no coherente", puesto que este último término significa efectuarla sin información de la fase portadora. En lugar de eso, empleamos una técnica "pseudo PSK" conocida como *corrimiento de fase por llaveo diferencial* (DPSK), que, en sentido general se considera como la forma no coherente de la PSK. En la práctica, la FSK M-aria y la DPSK M-aria son las formas de la modulación digital que se usan por lo común en sistemas no coherentes.

■ PROBABILIDAD DE ERROR

Una meta fundamental en los sistemas de transmisión de datos pasobanda corresponde al diseño óptimo del receptor, de modo que se minimice la probabilidad promedio del error de símbolo ante la presencia de *ruido blanco gaussiano aditivo* (AWGN). Con esta meta en mente, buena parte del material presentado en este capítulo se basa en las herramientas del análisis del espacio de señales que se presentó en el capítulo 5. De manera específica, en el estudio de cada sistema empezamos con la formulación de una constelación de señales y la construcción de regiones de decisión de acuerdo con la detección de señal de máxima verosimilitud sobre un canal AWGN. Estas formulaciones crean el escenario para evaluar la probabilidad del error de símbolo P_e. Dependiendo del método de modulación digital que se estudie, la evaluación de P_e procede de una de dos maneras:

- En el caso de ciertos métodos simples, como la PSK binaria coherente y la FSK binaria coherente, se obtienen fórmulas exactas para P_e.

- En el caso de un método más elaborado, como la PSK M-aria coherente y la FSK M-aria coherente, recurrimos al uso de la cota de unión para obtener una fórmula aproximada relativa a P_e.

Espectro de potencia

Para apreciar por completo las virtudes prácticas de los diferentes métodos de la modulación digital, necesitamos estudiar también el espectro de potencia de las señales moduladas resultantes. Este último aspecto es en particular importante en dos contextos: la ocupación del ancho de banda del canal y la interferencia de cocanal en sistemas multiplexados.

Dada una señal modulada $s(t)$, podemos describirla en términos de sus componentes en fase y en cuadratura como

$$\begin{aligned} s(t) &= s_I(t)\cos(2\pi f_c T) - s_Q(t)\,\mathrm{sen}(2\pi f_c t) \\ &= \mathrm{Re}\left[\tilde{s}(t)\exp(j2\pi f_c t)\right] \end{aligned} \tag{6.1}$$

donde Re[·] es la parte real de la expresión contenida en los paréntesis cuadrados. También tenemos

$$\tilde{s}(t) = s_I(t) + js_Q(t) \tag{6.2}$$

y

$$\exp(j2\pi f_c T) = \cos(2\pi f_c t) + j\,\mathrm{sen}(2\pi f_c t) \tag{6.3}$$

La señal $\tilde{s}(t)$ es la *envolvente compleja* (es decir, la versión de banda base) de la señal modulada (pasobanda) $s(t)$. Las componentes $s_I(t)$ y $s_Q(t)$, y por tanto $\tilde{s}(t)$, son todas señales pasobajas. Se definen de manera única en términos de la señal pasobanda $s(t)$ y de la frecuencia de portadora f_c, siempre que el ancho de banda medio de $s(t)$ sea menor que la frecuencia portadora f_c.

Sea $S_B(f)$ la densidad espectral de potencia de la envolvente compleja $\tilde{s}(t)$. Nos referimos a $S_B(f)$ como la *densidad espectral de potencia de banda base*. La densidad espectral de potencia, $S_S(f)$, de la señal pasobanda original $s(t)$ es una versión recorrida en frecuencia de $S_B(f)$, excepto por un factor de escalamiento, según se indica mediante

$$S_S(t) = \frac{1}{4}\left[S_B(f - f_c) + S_B(f + f_c)\right] \tag{6.4}$$

Por tanto, resulta suficiente evaluar la densidad espectral de potencia de banda base $S_B(f)$. Puesto que $\tilde{s}(t)$ es una señal pasobajas, el cálculo de $S_B(f)$ debe ser más simple que el de $S_S(f)$. (Vea el ejemplo 1.7.)

Eficiencia del ancho de banda

A lo largo de este libro hemos subrayado que el *ancho de banda del canal* y la *potencia transmitida* conforman dos "recursos de comunicación" fundamentales, cuya utilización eficiente proporciona el estímulo para buscar esquemas *espectralmente eficientes*. El objetivo principal de la modulación espectralmente eficiente es maximizar la eficiencia del ancho de banda definida como la *proporción entre la velocidad de datos en bits por segundo y el ancho de banda del canal utilizado efectivamente*. Un objetivo secundario consiste en alcanzar esta eficiencia del ancho de banda a un gasto prácticamente mínimo de potencia promedio de la señal o, equivalentemente, en un canal perturbado por ruido blanco gaussiano aditivo, a un gasto prácticamente mínimo de la relación señal a ruido promedio.

Con la velocidad de datos denotada por R_b y el ancho de banda del canal utilizado de manera efectiva por B, es posible expresar la eficiencia del ancho de banda, ρ, como

$$\rho = \frac{R_b}{B}\ \text{bits/s/Hz} \tag{6.5}$$

Recuerde del capítulo 4 que la eficiencia del ancho de banda es el producto de dos factores independientes: uno debido al posible uso de *codificación de multinivel* y el otro por la *formación espectral.* En la codificación de multinivel, la transmisión de información por el canal se realiza con base en bloques de bits en lugar de bits individuales. Con la formación espectral eficiente, el requerimiento de ancho de banda sobre el canal se reduce mediante el uso de filtros de formación de pulso que uniforma las transiciones abruptas en la forma de onda transmitida. Por sí solos estos dos factores son, en consecuencia, modos individuales importantes para determinar la eficiencia del ancho de banda de un sistema de transmisión de datos pasobanda de interés.

6.2 *Modelo de transmisión pasobanda*

En un sentido funcional, es posible modelar un sistema de transmisión de datos pasobanda como se indica en la figura 6.2. Primero, se supone que existe ahí una *fuente de mensajes* que emite un *símbolo* cada T segundos, con éstos pertenecientes a un alfabeto de M símbolos, que denotamos mediante $m_1, m_2, \ldots, m_M$. Las *probabilidades a priori* $P(m_1), P(m_2), \ldots, P(m_M)$ especifican la salida de la fuente de mensajes. Cuando los símbolos M del alfabeto son *igualmente probables,* escribimos

$$\begin{aligned} p_i &= P(m_i) \\ &= \frac{1}{M} \qquad \text{para toda } i \end{aligned} \tag{6.6}$$

La salida M-aria de la fuente de mensaje se presenta a un *codificador de transmisión de señal,* generándose un vector $\mathbf{s}_i$ correspondiente integrado por N elementos reales, correspondiendo uno de tales conjuntos a cada uno de los M símbolos del alfabeto de la fuente; la dimensión N es menor o igual que M. Con el vector $\mathbf{s}_i$ como entrada, el *modulador* construye entonces una señal *distinta* $s_i(t)$ de duración T segundos como la representación del símbolo m_i generado por la fuente de mensajes. La señal $s_i(t)$ es necesariamente una señal de energía, lo que se indica como

$$E_i = \int_0^T s_i^2(t)dt, \quad i = 1, 2, \ldots, M \tag{6.7}$$

Note que $s_i(t)$ es de valor real. Una señal de estas características se transmite cada T segundos. La señal particular que se elige para la transmisión depende de algún modo del mensaje entrante y posiblemente de las señales transmitidas en los periodos anteriores. Con una portadora senoidal, la característica que utiliza el modulador para distinguir una señal de otra es un cambio de *escalón* en la amplitud, frecuencia o fase portadora. (Algunas veces se usa una forma híbrida de modulación que combina cambios tanto de amplitud como de fase o de amplitud y de frecuencia.)

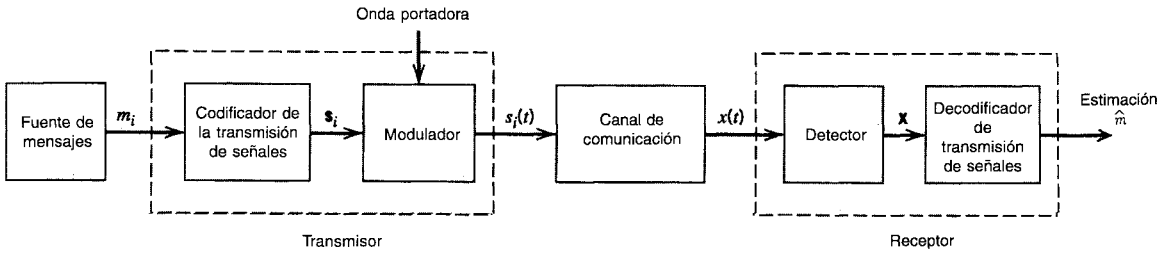


FIGURA 6.2 Modelo funcional del sistema de transmisión de datos pasobanda.

Volviendo al modelo funcional de la figura 6.2, el canal de comunicación pasobanda, que acopla al transmisor con el receptor, se supone que tiene dos características:

1. El canal es lineal, con un ancho de banda lo suficientemente amplio para alojar la transmisión de la señal modulada $s_i(t)$ con distorsión imperceptible o sin ella.
2. El ruido del canal $w(t)$ es la función de muestreo del proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$.

Las suposiciones hechas aquí son básicamente las mismas a las que se recurrió en el capítulo 5 al tratar con el análisis espacial de señales.

El receptor, que consiste en un *detector* seguido por un *decodificador de transmisión de señal*, realiza dos funciones:

1. Invierte las operaciones realizadas en el transmisor.
2. Minimiza el efecto del ruido del canal sobre la estimación $\hat{m}$ calculada para el símbolo transmitido m_i.

6.3 Corrimiento de fase por llaveo coherente

Al contar con el material básico sobre la detección coherente de señales en ruido blanco gaussiano aditivo que presentamos en el capítulo 5, estamos listos para estudiar los sistemas de transmisión de datos pasobanda específicos. En esta sección nos enfocamos en el corrimiento de fase por llaveo coherente (PSK), considerando la PSK binaria, la QPSK y sus variantes, y terminando con la PSK M-aria.

Corrimiento de fase por llaveo binario

En un sistema PSK binario coherente, el par de señales $s_1(t)$ y $s_2(t)$ utilizado para representar los símbolos binarios 1 y 0, respectivamente, está definido por

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \tag{6.8}$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \tag{6.9}$$

donde $0 \le t \le T_b$, y E_b es la *energía por bit de la señal transmitida*. Para asegurar que cada bit transmitido contiene un número integral de ciclos de la onda portadora, la frecuencia de portadora f_c se elige igual a n_c/T_b para algún entero fijo n_c. Un par de ondas senoidales que difieren sólo en un corrimiento de fase relativo de 180 grados, como se define en las ecuaciones (6.8) y (6.9), recibe el nombre de *señales antípodas*.

A partir de este par de ecuaciones resulta claro que, en el caso de la PSK binaria, sólo hay una función básica de energía unitaria, a saber,

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t < T_b \tag{6.10}$$

En ese caso podemos expresar las señales transmitidas $s_1(t)$ y $s_2(t)$ en términos de $\phi_1(t)$ como sigue:

$$s_1(t) = \sqrt{E_b}\phi_1(t), \qquad 0 \le t < T_b \tag{6.11}$$

y

$$s_2(t) = -\sqrt{E_b}\phi_1(t), \qquad 0 \le t < T_b \tag{6.12}$$

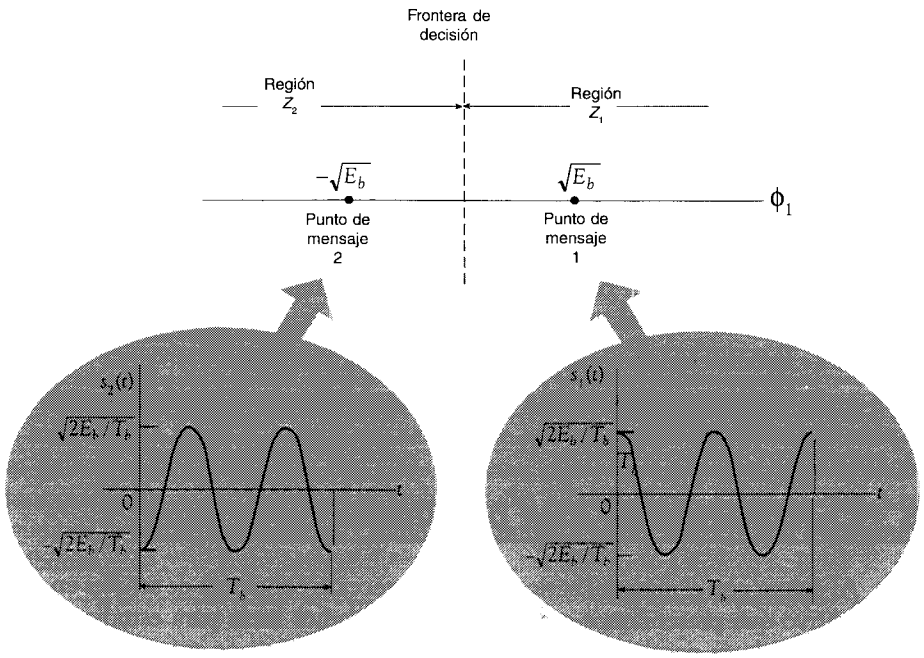


FIGURA 6.3 Diagrama del espacio de señales para un sistema PSK binario coherente. Las formas de onda describen las señales transmitidas $s_1(t)$ y $s_2(t)$, desplegadas en los insertos, suponiendo $n_c = 2$.

Un sistema PSK binario coherente está caracterizado por tener un espacio de señal que es unidimensional (es decir, $N = 1$), con una constelación consistente en dos puntos de mensaje (es decir, $M = 2$). Las coordenadas de los puntos de mensaje son

$$\begin{aligned} s_{11} &= \int_0^{T_b} s_1(t)\phi_1(t)\,dt \\ &= +\sqrt{E_b} \end{aligned} \tag{6.13}$$

y

$$\begin{aligned} s_{21} &= \int_0^{T_b} s_2(t)\phi_1(t)\,dt \\ &= -\sqrt{E_b} \end{aligned} \tag{6.14}$$

Los puntos de mensaje correspondientes a $s_1(t)$ se localizan en $s_{11} = +\sqrt{E_b}$; y el punto de mensaje correspondiente a $s_2(t)$ se ubica en $s_{21} = -\sqrt{E_b}$. La figura 6.3 presenta el diagrama del espacio de señales para la PSK binaria. Esta figura incluye también dos insertos, que muestran formas de onda de ejemplo de señales antípodas que representan a $s_1(t)$ y $s_2(t)$. Advierta que la constelación de la figura 6.3 tiene una energía promedio mínima.

Probabilidad de error de la PSK binaria

Para realizar una *regla que tome una decisión* en favor del símbolo 1 o el símbolo 0, aplicamos la ecuación (5.59) del capítulo 5. De manera específica, dividimos el espacio de señales de la figura 6.3 en dos regiones:

- El conjunto de puntos más cercano al punto de mensaje 1 en $+\sqrt{E_b}$.
- El conjunto de puntos más cercano al punto de mensaje 2 en $-\sqrt{E_b}$.

Esto se lleva a cabo construyendo el punto medio de la línea que une estos dos puntos de mensaje y marcando después las regiones de decisión apropiadas. En la figura 6.3 estas regiones de decisión se marcan como Z_1 y Z_2, de acuerdo con el punto de mensaje alrededor del cual se construyen.

La regla de decisión implica ahora decidir simplemente qué señal $s_1(t)$ (es decir, el símbolo binario 1) se transmitió si el punto de la señal recibida cae en la región Z_1, y decidir qué señal $s_2(t)$ (es decir, el símbolo binario 0) se transmitió si la señal recibida cae en la región Z_2. Sin embargo, es posible realizar dos tipos de decisiones erróneas. La señal $s_2(t)$ se transmite, pero el ruido es tal que el punto de la señal recibida cae dentro de la región Z_1 y, por ello, el receptor decide en favor de la señal $s_1(t)$. De manera alternativa, la señal $s_1(t)$ se transmite, pero el ruido es tal que el punto de la señal que se recibe cae dentro de la región Z_2 y, por esa causa, el receptor decide en favor de la señal $s_2(t)$.

Para calcular la probabilidad de efectuar un error del primer tipo, advertimos de la figura 6.3 que la región de decisión asociada con el símbolo 1 o señal $s_1(t)$ está descrita por medio de

$$Z_1 : 0 < x_1 < \infty$$

donde el elemento observable x_1 se relaciona con la señal recibida $x(t)$ mediante

$$x_1 = \int_0^{T_b} x(t)\phi_1(t)\,dt \tag{6.15}$$

La función de densidad de probabilidad condicional de una variable aleatoria X_1, debido a que se transmitió el símbolo 0 [es decir, la señal es $s_2(t)$], está definida por

$$\begin{aligned} f_{X_1}(x_1|0) &= \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_1 - s_{21})^2\right] \\ &= \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{1}{N_0}(x_1 + \sqrt{E_b})^2\right] \end{aligned} \tag{6.16}$$

La probabilidad condicional de que el receptor decida en favor del símbolo 1, debido a que se transmitió el símbolo 0, es entonces

$$\begin{aligned} p_{10} &= \int_0^\infty f_{X_1}(x_1|0)\,dx_1 \\ &= \frac{1}{\sqrt{\pi N_0}} \int_0^\infty \exp\left[-\frac{1}{N_0}(x_1 + \sqrt{E_b})^2\right] dx_1 \end{aligned} \tag{6.17}$$

Haciendo

$$z = \frac{1}{\sqrt{N_0}}(x_1 + \sqrt{E_b}) \tag{6.18}$$

y cambiando la variable de integración de x_1 a z, es posible reescribir la ecuación (6.17) en la forma compacta

$$\begin{aligned} p_{10} &= \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/N_0}}^\infty \exp(-z^2)\,dz \\ &= \frac{1}{2}\operatorname{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \end{aligned} \tag{6.19}$$

donde ferc$(\cdot)$ es la función de error complementaria.

Considere a continuación un error del segundo tipo. Advertimos que el espacio de señales de la figura 6.3 es simétrico con respecto al origen. Por tanto, se concluye que p_{01}, la probabilidad condicional de que el receptor decida en favor del símbolo 0, debido a que se transmitió el símbolo 1, tiene también el mismo valor que en la ecuación (6.19).

De esa manera, al promediar las probabilidades de error condicional p_{10} y p_{01}, encontramos que la *probabilidad promedio del error de símbolo* o, equivalentemente, la *tasa de error de bits para la PSK binaria coherente* (suponiendo símbolos igualmente probables) es

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.20}$$

Cuando aumentamos la energía por bit de la señal transmitida, E_b, para una densidad espectral de ruido especificada N_0, se apartan los puntos de mensaje correspondientes a los símbolos 1 y 0, y la probabilidad promedio de error P_e se reduce en forma correspondiente de acuerdo con la ecuación (6.20), la cual se satisface en forma intuitiva.

Generación y detección de señales PSK binarias coherentes

Para generar una señal PSK binaria, vemos de las ecuaciones (6.8)-(6.10) que debemos representar la secuencia binaria de entrada en forma polar con los símbolos 1 y 0 indicados por niveles de amplitud constante de $+\sqrt{E_b}$ y $-\sqrt{E_b}$, respectivamente. Esta codificación de transmisión de señal se realiza mediante un codificador de nivel polar sin retorno a cero (NRZ). La onda binaria que resulta y la portadora senoidal $\phi_1(t)$, cuya frecuencia es $f_c = (n_c/T_b)$ para algún entero fijo n_c, se aplican a un modulador de producto, como en la figura 6.4*a*. La portadora y los pulsos de temporización utilizados para generar la onda binaria suelen extraerse de un reloj maestro común. La onda PSK deseada se obtiene a la salida del modulador.

Para detectar la secuencia binaria original de unos y ceros, aplicamos la señal PSK ruidosa $x(t)$ (a la salida del canal) en un correlator, el cual se alimenta también con una señal de referencia coherente $\phi_1(t)$ generada en forma local, como en la figura 6.4*b*. La salida del correlator, x_1, se compara con un umbral de cero volts. Si $x_1 > 0$, el receptor decide en favor del símbolo 1.

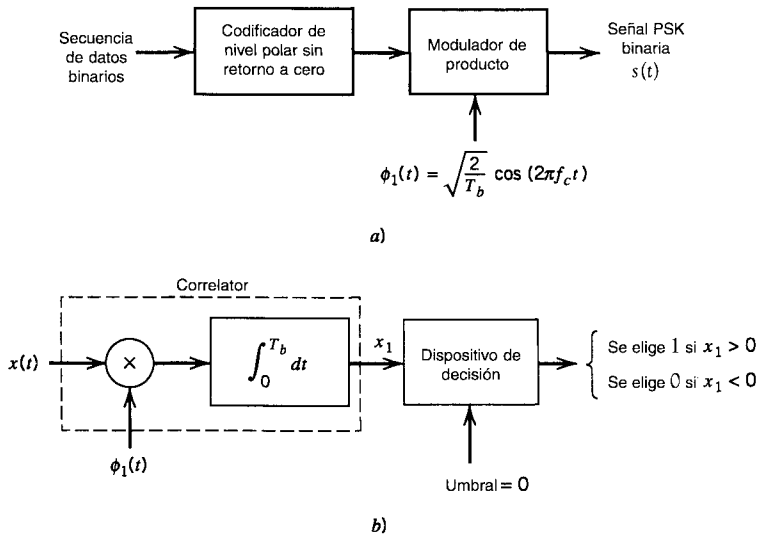


FIGURA 6.4 Diagrama de bloques para *a*) un transmisor PSK binario y *b*) receptor PSK binario coherente.

Por otra parte, si $x_1 < 0$, decide en favor del símbolo 0. Si x_1 es exactamente cero, el receptor realiza una suposición aleatoria en favor de 0 o 1.

Espectro de potencia de señales PSK binarias

De acuerdo con el modulador de la figura 6.4*a*, vemos que la envolvente compleja de una onda PSK binaria está compuesta únicamente por la componente de fase. Además, dependiendo de si tenemos el número 1 o el símbolo 0 a la entrada del modulador durante el intervalo de transmisión de señales $0 \le t \le T_b$, encontramos que esta componente en fase es igual a $+g(t)$ o $-g(t)$, respectivamente, donde $g(t)$ es la *función de formación del símbolo* definida por

$$g(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}}, & 0 \le t \le T_b \\ 0, & \text{en otro caso} \end{cases} \tag{6.21}$$

Suponemos que la onda binaria de entrada es aleatoria, con símbolos 1 y 0 igualmente probables, y los símbolos se transmiten durante los diferentes periodos en los que son estadísticamente independientes. En el ejemplo 1.6 del capítulo 1 se demuestra que la densidad espectral de potencia de una onda binaria aleatoria descrita de esa manera es igual a la densidad espectral de energía de la función de formación de símbolo dividida entre la duración de este último. La densidad espectral de energía de una señal transformable de Fourier $g(t)$ se define como la magnitud al cuadrado de la transformada de Fourier de la señal. Por tanto, la densidad espectral de potencia de banda base de una señal PSK binaria es igual a

$$\begin{aligned} S_B(f) &= \frac{2E_b \operatorname{sen}^2(\pi T_b f)}{(\pi T_b f)^2} \\ &= 2E_b \operatorname{senc}^2(T_b f) \end{aligned} \tag{6.22}$$

Este espectro de potencia disminuye como el cuadrado inverso de la frecuencia, en la forma en que se indica en la figura 6.5.

La figura 6.5 incluye también una gráfica de la densidad espectral de potencia de banda base de una señal FSK binaria, y los detalles de esta misma se presentan en la sección 6.5. La comparación de estos dos espectros se deja para esa sección.

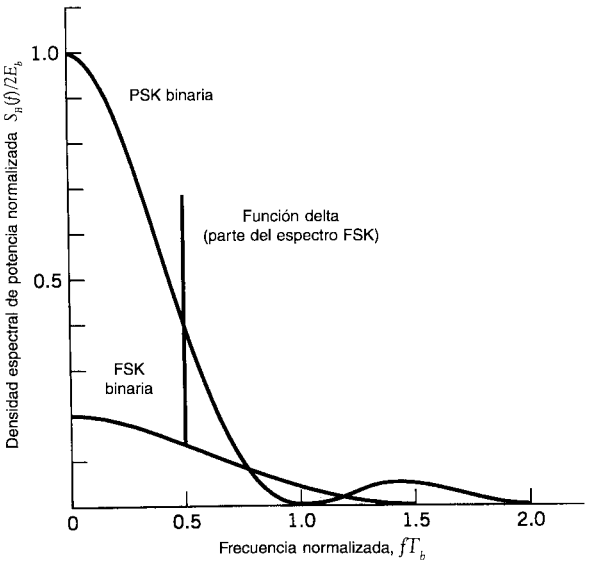


FIGURA 6.5 Espectros de potencia de señales PSK y FSK binarias.

■ CORRIMIENTO DE CUADRIFASE POR LLAVEO

La obtención de un desempeño confiable, ejemplificado por una muy baja probabilidad de error, es una meta importante en el diseño de un sistema de comunicación digital. Otra meta importante corresponde a la utilización eficiente del ancho de banda del canal. En esta subsección estudiamos un esquema de modulación conservando el ancho de banda, que se conoce como corrimiento de cuadrifase por llaveo coherente, el cual es un ejemplo del *multiplexado de portadora en cuadratura.*

En el *corrimiento en cuadrifase por llaveo* (QPSK), como con la PSK binaria, la información que porta la señal transmitida está contenida en la fase. En particular, la fase portadora toma uno de cuatro valores igualmente espaciados, tales como $\pi/4$, $3\pi/4$, $5\pi/4$ y $7\pi/4$. Para este conjunto de valores es posible definir la señal transmitida como

$$s_i(t) = \begin{cases} \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right], & 0 \le t \le T \\ 0, & \text{en otro lado} \end{cases} \tag{6.23}$$

donde $i = 1, 2, 3, 4$; E es la energía por símbolo de la señal transmitida, y T es la duración del símbolo. La frecuencia f_c de la portadora es igual a n_c/T para algún entero fijo n_c. Cada valor posible de la fase corresponde a un dibit único. Así, por ejemplo, es posible elegir el siguiente conjunto de valores de fase para representar el conjunto de dibits de *codificación Gray* 10, 00, 01 y 11, donde solamente un único bit se cambia de un dibit al siguiente.

Diagrama del espacio de señales de la QPSK

Utilizando una identidad trigonométrica bien conocida, es posible utilizar la ecuación (6.23) para redefinir la señal transmitida $s_i(t)$ correspondiente al intervalo $0 \le t \le T$ en la forma equivalente:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[(2i-1)\frac{\pi}{4}\right] \cos(2\pi f_c t) - \sqrt{\frac{2E}{T}} \operatorname{sen}\left[(2i-1)\frac{\pi}{4}\right] \operatorname{sen}(2\pi f_c t) \tag{6.24}$$

donde $i = 1, 2, 3, 4$. Con base en esta representación, podemos plantear las siguientes observaciones:

- Existen dos funciones básicas ortogonales, $\phi_1(t)$ y $\phi_2(t)$, contenidas en la expansión de $s_i(t)$. De manera específica, $\phi_1(t)$ y $\phi_2(t)$ se definen mediante un par de *portadoras en cuadratura:*

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t), \qquad 0 \le t \le T \tag{6.25}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}} \operatorname{sen}(2\pi f_c t), \qquad 0 \le t \le T \tag{6.26}$$

TABLA 6.1 *Caracterización del espacio de señales de la QPSK*

Entrada de dibit con codificación Gray	*Fase de la señal QPSK (radianes)*	*Coordenadas de los puntos de mensaje*	
		s_{i1}	s_{i2}
10	$\pi/4$	$+\sqrt{E/2}$	$-\sqrt{E/2}$
00	$3\pi/4$	$-\sqrt{E/2}$	$-\sqrt{E/2}$
01	$5\pi/4$	$-\sqrt{E/2}$	$+\sqrt{E/2}$
11	$7\pi/4$	$+\sqrt{E/2}$	$+\sqrt{E/2}$

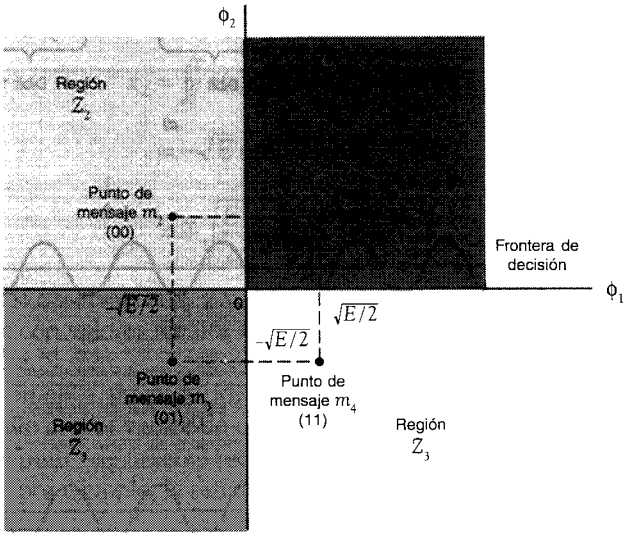


FIGURA 6.6 Diagrama del espacio de señales del sistema QPSK coherente.

- Existen cuatro puntos de mensaje, y los vectores de señal asociados se definen por medio de

$$\mathbf{s}_i = \begin{bmatrix} \sqrt{E}\cos\left((2i-1)\dfrac{\pi}{4}\right) \\ -\sqrt{E}\,\text{sen}\left((2i-1)\dfrac{\pi}{4}\right) \end{bmatrix}, \quad i = 1, 2, 3, 4 \tag{6.27}$$

  Los elementos de los vectores de señal, a saber, s_{i1} y s_{i2}, tienen sus valores resumidos en la tabla 6.1. Las primeras dos columnas proporcionan el dibit y la fase asociados de la señal QPSK.

En consecuencia, una señal QPSK tiene una constelación de señales bidimensional ($N = 2$) y cuatro puntos de mensaje ($M = 4$), cuyos ángulos de fase aumentan en dirección contraria a la de las manecillas del reloj, según se ilustra en la figura 6.6. Como en el caso de la PSK binaria, la señal QPSK tiene una energía promedio mínima.

EJEMPLO 6.1

La figura 6.7 ilustra las secuencias y formas de onda implicadas en la generación de una señal QPSK. La secuencia binaria de entrada 01101000 se muestra en la figura 6.7*a*. Ésta se divide en otras dos secuencias, consistentes en bits de número impar y par de la secuencia de entrada. Estas dos secuencias se muestran en las líneas superiores de las figuras 6.7*b* y 6.7*c*. Las formas de onda que representan a los dos componentes de la señal QPSK, es decir, $s_{i1}\phi_1(t)$ y $s_{i2}\phi_2(t)$, también se presentan en las figuras 6.7*b* y 6.7*c*, respectivamente. Estas dos formas de onda pueden considerarse individualmente como ejemplos de una señal PSK binaria. Al sumarlas obtenemos la forma de onda QPSK que se muestra en la figura 6.7*d*.

Para definir la regla de decisión correspondiente a la detección de la secuencia de datos transmitida, dividimos el espacio de señales en cuatro regiones, de acuerdo con la ecuación (5.59) del capítulo 5. Las regiones individuales se definen mediante el conjunto de puntos más cercanos al punto de mensaje representado por los vectores de señal $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ y $\mathbf{s}_4$. Esto se consigue sin dificultad construyendo los bisectores perpendiculares del cuadrado formado al unir los cuatro puntos del mensaje y al

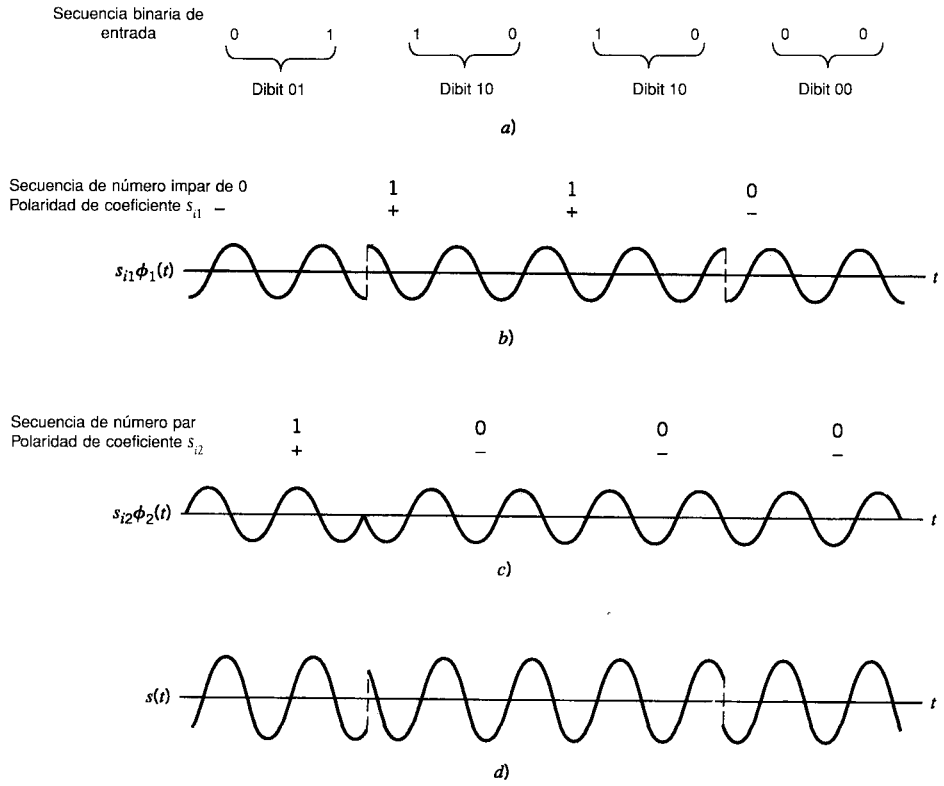


FIGURA 6.7 *a*) Secuencia binaria de entrada. *b*) Número impar de bits de la secuencia de entrada y onda PSK binaria asociada. c) Número par de bits de la secuencia de entrada y onda PSK binaria asociada. *d*) Forma de onda QPSK definida como $s(t) = s_{i1}\phi_1(t) + s_{i2}\phi_2(t)$.

marcar después las regiones apropiadas. De este modo encontramos que las regiones de decisión son cuadrantes cuyos vértices coinciden con el origen. Estas regiones se indican mediante Z_1, Z_2, Z_3 y Z_4 en la figura 6.6, de acuerdo con los puntos de mensaje alrededor de los cuales se construyen.

Probabilidad de error de la QPSK

En un sistema QPSK coherente, la señal recibida $x(t)$ se define por medio de

$$x(t) = s_i(t) + w(t), \qquad \begin{cases} 0 \le t \le T \\ i = 1, 2, 3, 4 \end{cases} \tag{6.28}$$

donde $w(t)$ es la función de muestreo de un proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$. De manera correspondiente, el vector de observación $\mathbf{x}$ tiene dos elementos, x_1 y x_2, definidos por

$$\begin{aligned} x_1 &= \int_0^T x(t)\phi_1(t)\,dt \\ &= \sqrt{E}\cos\left[(2i-1)\frac{\pi}{4}\right] + w_1 \\ &= \pm\sqrt{\frac{E}{2}} + w_1 \end{aligned} \tag{6.29}$$

y

$$\begin{aligned} x_2 &= \int_0^T x(t)\phi_2(t)\,dt \\ &= -\sqrt{E}\,\text{sen}\left[(2i-1)\frac{\pi}{4}\right] + w_2 \\ &= \mp\sqrt{\frac{E}{2}} + w_2 \end{aligned} \tag{6.30}$$

De ese modo los elementos observables x_1 y x_2 son valores de muestreo de variables aleatorias gaussianas independientes con valores medios iguales a $\pm\sqrt{E/2}$ y $\mp\sqrt{E/2}$, respectivamente, y con varianza común igual a $N_0/2$.

La regla de decisión implica ahora decidir simplemente que $s_1(t)$ se transmitió si el punto de la señal recibida asociado con el vector de observación $\mathbf{x}$ cae dentro de la región Z_1, decidir que $s_2(t)$ se transmitió si el punto de la señal recibida cae dentro de la región Z_2, etc. Se efectuará una decisión errónea si, por ejemplo, la señal $s_4(t)$ se transmite pero el ruido $w(t)$ es tal que el punto de la señal recibida cae fuera de la región Z_4.

Para calcular la probabilidad promedio del error de símbolo, advertimos con base en la ecuación (6.24) que el sistema QPSK coherente es de hecho equivalente a los dos sistemas PSK binarios coherentes que trabajan en paralelo y que usan dos portadoras que están en cuadratura de fase. Esto es únicamente un enunciado de la propiedad de multiplexado de portadora en cuadratura de la QPSK coherente. La salida del canal en fase x_1 y la salida del canal en cuadratura x_2 (es decir, los dos elementos del vector de observación $\mathbf{x}$) pueden verse como salidas individuales de los dos sistemas PSK binarios coherentes. Por tanto, de acuerdo con las ecuaciones (6.29) y (6.30), estos dos sistemas PSK binarios pueden caracterizarse de la manera que sigue:

- La energía de la señal por bit es $E/2$.
- La densidad espectral de ruido es $N_0/2$.

En consecuencia, utilizando la ecuación (6.20) para la probabilidad promedio de un error de bit de un sistema PSK binario coherente, podemos establecer ahora que la probabilidad promedio de error de bit en *cada* canal del sistema QPSK coherente es

$$\begin{aligned} P' &= \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E/2}{N_0}}\right) \\ &= \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right) \end{aligned} \tag{6.31}$$

Otro punto importante a observar es que los errores de bit en los canales en fase y en cuadratura del sistema QPSK coherente son estadísticamente independientes. El canal en fase toma una decisión de uno de los dos bits que constituyen un símbolo (dibit) de la señal QPSK, y el canal en cuadratura se ocupa del otro bit. En consecuencia, la *probabilidad promedio de una decisión correcta* que resulta de la acción combinada de los dos canales trabajando en conjunto es

$$\begin{aligned} P_c &= (1-P')^2 \\ &= \left[1 - \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right)\right]^2 \\ &= 1 - \text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right) + \frac{1}{4}\text{ferc}^2\left(\sqrt{\frac{E}{2N_0}}\right) \end{aligned} \tag{6.32}$$

La probabilidad promedio del error de símbolo para la QPSK coherente es entonces

$$\begin{aligned} P_e &= 1 - P_c \\ &= \text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right) - \frac{1}{4}\text{ferc}^2\left(\sqrt{\frac{E}{2N_0}}\right) \end{aligned} \tag{6.33}$$

En la región donde $(E/2N_0) >> 1$, es posible ignorar el término cuadrático en el segundo miembro de la ecuación (6.33), por lo que aproximamos la fórmula para la probabilidad promedio del error de símbolo relativa a la QPSK coherente como

$$P_e \simeq \text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right) \tag{6.34}$$

La fórmula de la ecuación (6.34) puede obtenerse también de otro modo ilustrativo, utilizando el diagrama del espacio de señales de la figura 6.6. Puesto que los cuatro puntos de mensaje de este diagrama son simétricos circularmente con respecto al origen, podemos aplicar la ecuación (5.92), que se reproduce en la forma

$$P_e \leq \frac{1}{2}\sum_{\substack{k=1 \\ k\neq i}}^{4}\text{ferc}\left(\frac{d_{ik}}{2\sqrt{N_0}}\right) \quad \text{para toda } i \tag{6.35}$$

Considere, por ejemplo, el punto de mensaje m_1 (correspondiente al dibit 10) elegido como el punto de mensaje transmitido. Los puntos de mensaje m_2 y m_4 (correspondientes a los dibits 00 y 11) son los más cercanos a m_1. A partir de la figura 6.6 encontramos de inmediato que m_1 equidista de m_2 y m_4 en un sentido euclidiano, como indica

$$d_{12} = d_{14} = \sqrt{2E}$$

Suponiendo que E/N_0 es suficientemente grande para ignorar la contribución de la mayor parte de los puntos de mensaje distantes m_3 (correspondientes al dibit 01) relativos a m_1, encontramos que el uso de la ecuación (6.35) produce una expresión aproximada para P_e que es la misma que la ecuación (6.34). Advierta que al confundir ya sea m_2 o m_4 por m_1, se comete un solo error de bit. Por otra parte, al confundir m_3 por m_1, se cometen dos errores de bit. Para un E/N_0 suficientemente elevado, la probabilidad de que ambos bits de un símbolo sean erróneos es mucho menor que la correspondiente a un solo bit, lo cual es una justificación adicional para ignorar m_3 al calcular P_e cuando se envía m_1.

En un sistema QPSK advertimos que, puesto que hay dos bits por símbolo, la energía de la señal transmitida por símbolo es el doble de la energía de la señal por bit, de acuerdo con

$$E = 2E_b \tag{6.36}$$

Es por eso que al expresar la probabilidad promedio del símbolo en términos de la proporción E_b/N_0, puede escribirse

$$P_e \simeq \text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.37}$$

Con la codificación Gray utilizada para los símbolos entrantes encontramos, de acuerdo con las ecuaciones (6.31) y (6.36), que la *tasa de error de bits* de QPSK es exactamente

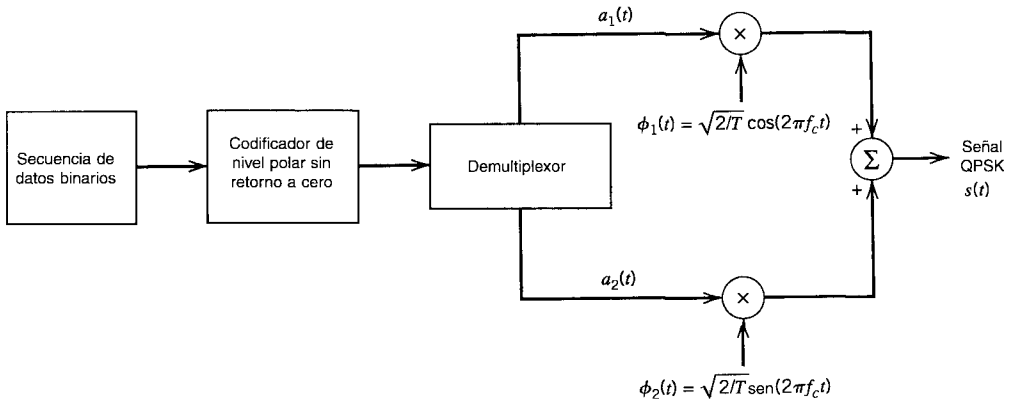
$$\text{BER} = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.38}$$

Por tanto, resulta viable establecer que un sistema QPSK coherente alcanza la misma probabilidad promedio de error de bits que un sistema PSK binario coherente para la misma tasa de bits y la misma E_b/N_0, pero que sólo usa la mitad del ancho de banda del canal. Planteado de diferente manera, para la misma E_b/N_0 y consecuentemente la misma probabilidad promedio de error de bits, un sistema QPSK coherente transmite información al doble de la tasa de bits de un sistema

PSK coherente para el mismo ancho de banda del canal. Con respecto a un desempeño preestablecido, la QPSK utiliza un ancho de banda de canal mejor que la PSK binaria, lo cual explica que se prefiera en la práctica a la QPSK sobre la PSK binaria.

Generación y detección de señales QPSK coherentes

Considere a continuación la generación y detección de señales QPSK. La figura 6.8*a* muestra un diagrama de bloques de un transmisor QPSK característico. La secuencia de datos binarios entrante se transforma primero en una forma polar por medio de un codificador *de nivel sin retorno a cero*. De tal modo, los símbolos 1 y 0 se representan por medio de $+\sqrt{E_b}$ y $-\sqrt{E_b}$, respectivamente. Esta onda binaria se divide en seguida empleando un *demultiplexor* en dos ondas binarias independientes compuestas de bits de entrada de número par e impar, denotadas por $a_1(t)$ y $a_2(t)$. Observamos que en cualquier intervalo de transmisión de señales, la amplitud de $a_1(t)$ y $a_2(t)$ es igual a s_{i1} y s_{i2}, respectivamente, dependiendo del dibit particular que se esté transmitiendo. Las dos ondas binarias $a_1(t)$ y $a_2(t)$ se utilizan para modular un par de portadoras en cuadratura o funciones básicas ortonormales: $\phi_1(t)$ igual a $\sqrt{2/T}\cos(2\pi f_c t)$ y $\phi_2(t)$ igual a $\sqrt{2/T}\,\text{sen}(2\pi f_c t)$. El



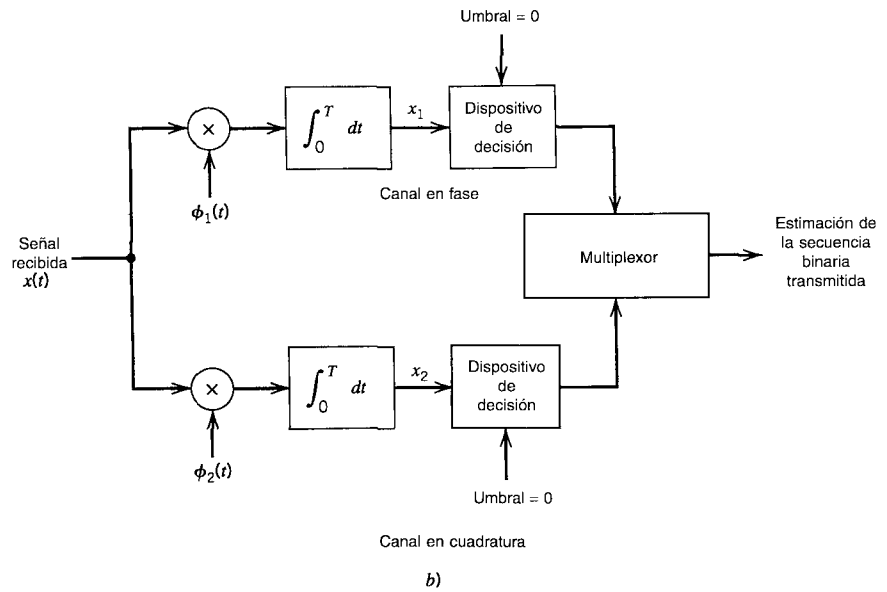


FIGURA 6.8 Diagrama de bloques de *a*) transmisor QPSK y *b*) receptor QPSK coherente.

resultado es un par de señales PSK binarias, que pueden detectarse de manera independiente debido a la ortogonalidad de $\phi_1(t)$ y $\phi_2(t)$. Por último, las dos señales PSK binarias se suman para producir la señal QPSK que se desea.

El receptor QPSK consiste en un par de correlatores con una entrada común y se alimenta con un par de señales de referencia $\phi_1(t)$ y $\phi_2(t)$ generado en forma local, como en la figura 6.8*b*. Las salidas x_1 y x_2 del correlator, producidas en respuesta a la señal recibida $x(t)$, se comparan cada una con el umbral de cero. Si $x_1 > 0$, se toma una decisión en favor del símbolo 1, para la salida del canal en fase, pero si $x_1 < 0$, se toma una decisión en favor del símbolo 0. Similarmente, si $x_2 > 0$, se decide a favor del símbolo 1 para la salida del canal en cuadratura, pero si $x_2 < 0$, la decisión es a favor del símbolo 0. Por último, estas dos secuencias binarias a las salidas de canal en fase y en cuadratura se combinan en *un multiplexor* para reproducir la secuencia binaria original a la entrada del transmisor con la probabilidad mínima de error del símbolo en un canal AWGN.

Espectros de potencia de señales QPSK

Suponga que la onda binaria es aleatoria a la entrada del modulador, siendo los símbolos 1 y 0 igualmente probables, y de manera que los símbolos transmitidos durante periodos adyacentes sean independientes estadísticamente. Hacemos las siguientes observaciones en relación con las componentes en fase y en cuadratura de una señal QPSK:

1. Dependiendo del dibit enviado durante el intervalo de la transmisión de la señal $-T_b \le t \le T_b$, la componente en fase es igual a $+g(t)$ o $-g(t)$, y similarmente para la componente en cuadratura. La $g(t)$ denota la función de formación del símbolo, definida por

$$g(t) = \begin{cases} \sqrt{\dfrac{E}{T}}, & 0 \le t \le T \\ 0, & \text{en otro caso} \end{cases} \tag{6.39}$$

Por tanto, las componentes en fase y en cuadratura tienen una densidad espectral de potencia común, a saber, $E\,\text{senc}^2(Tf)$.

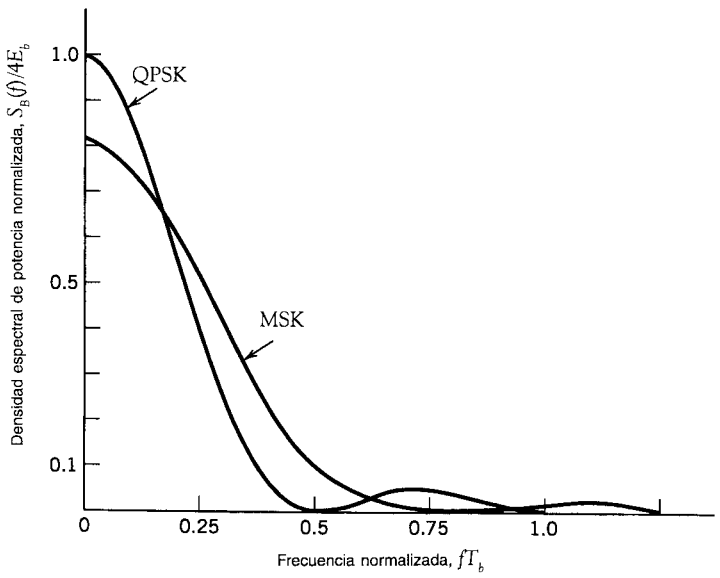


FIGURA 6.9 Espectros de potencia de señales QPSK y MSK.

2. Las componentes en fase y en cuadratura son estadísticamente independientes. Por tanto, la densidad espectral de potencia de banda base de la señal QPSK es igual a la suma de las densidades espectrales de potencia individuales de las componentes en fase y en cuadratura, por lo que es posible escribir

$$\begin{aligned} S_B(f) &= 2E\,\text{senc}^2(Tf) \\ &= 4E_b\,\text{senc}^2(2T_bf) \end{aligned} \tag{6.40}$$

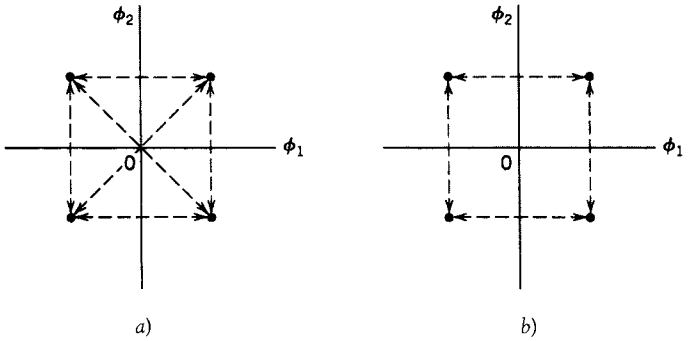
En la figura 6.9 se grafica $S_B(f)$, normalizada con respecto a $4E_b$, en función de la frecuencia normalizada fT_b. Esta figura incluye también una gráfica de la densidad espectral de potencia de banda base de una cierta forma de FSK binaria llamada corrimiento mínimo por llaveo, cuya evaluación se presenta en la sección 6.5. La comparación de estos dos ejemplos se posterga para esa sección.

QPSK DE COMPENSACIÓN

El diagrama del espacio de señales de la figura 6.10*a* comprende todas las posibles transiciones de fase que pueden surgir en la generación de una señal QPSK. De manera más específica, examinando la forma de onda QSPK que se ilustra en la figura 6.7 para el ejemplo 6.1, se establecen las siguientes observaciones:

1. La fase portadora cambia en ±180 grados siempre que cambien de signo tanto las componentes en fase como en cuadratura de la señal QPSK. Un ejemplo de esta situación se ilustra en la figura 6.7 cuando la secuencia binaria de entrada cambia del dibit 01 al dibit 10.
2. La fase portadora cambia en ±90 grados siempre que las componentes en fase o en cuadratura cambien de signo. Un ejemplo de esta segunda situación se ilustra en la figura 6.7 cuando la secuencia binaria de entrada varía del dibit 10 al dibit 00, durante lo cual cambia de signo la componente en fase, en tanto que queda invariable la componente en cuadratura.
3. La fase portadora permanece sin cambio cuando ni la componente en fase ni la componente en cuadratura cambian de signo. Esta última situación se ilustra en la figura 6.7 cuando el dibit 10 se transmite en dos intervalos de símbolo sucesivos.

La situación 1 y, a un grado mucho menor la situación 2, pueden ser de interés particular cuando la señal QPSK se filtra durante el curso de la transmisión, antes de la detección. Específicamente, los corrimientos de 180 y 90 grados en la fase portadora posiblemente produzcan cambios en la amplitud de la portadora (es decir, la envolvente de la señal QPSK) causando por ello errores de símbolo adicionales en la detección.



a) *b)*

FIGURA 6.10 Posibles trayectorias para la conmutación entre los puntos de mensaje en *a)* QPSK y *b)* QPSK de compensación.

El grado de las fluctuaciones de amplitud que exhiben las señales QPSK se reduce utilizando *QPSK de compensación*.[2] En esta variante de la QPSK, la secuencia de bits responsables para generar la componente en cuadratura está retardada (es decir, descompensada) por la mitad del intervalo de símbolo con respecto a la secuencia de bits responsables de la generación de la componente en fase. De modo específico, las dos funciones básicas de la QPSK de compensación se definen por medio de

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t), \qquad 0 \le t \le T \tag{6.41}$$

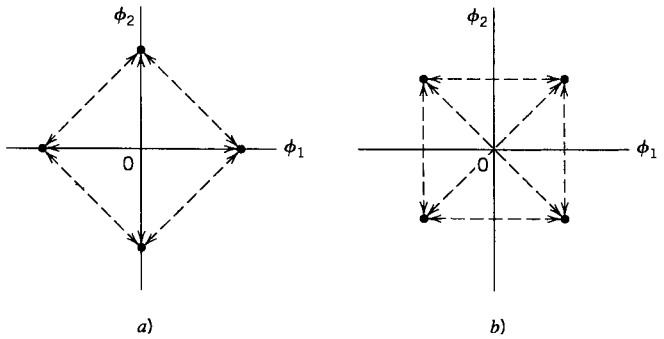
$$\phi_2(t) = \sqrt{\frac{2}{T}}\,\text{sen}(2\pi f_c t), \qquad \frac{T}{2} \le t \le \frac{3T}{2} \tag{6.42}$$

En consecuencia, a diferencia de la QPSK, las transiciones de fase que son probables de ocurrir en la QPSK de compensación se confinan a ±90 grados, como se indica en el diagrama del espacio de señales de la figura 6.10*b*. Sin embargo, transiciones de fase de ±90 grados en la QPSK de compensación ocurren dos veces más frecuentemente, aunque con la mitad de la intensidad que se encuentra en la QPSK. Puesto que, además de las transiciones de fase de ±90 grados, también ocurren transiciones de fase de ±180 grados en la QPSK, encontramos que las fluctuaciones de amplitud en la QPSK de compensación debidas al filtrado tienen una amplitud más pequeña que en el caso de la QPSK.

A pesar del retardo $T/2$ aplicado a la función básica $\phi_2(t)$ en la ecuación (6.42) comparada con la de la ecuación (6.26), la QPSK de compensación tiene exactamente la misma probabilidad del error de símbolo en un canal AWGN que la QPSK. La equivalencia del desempeño ante el ruido entre estos esquemas de corrimiento de fase por llaveo supone el uso de detección coherente. La razón de la equivalencia es que la independencia estadística de las componentes en fase y en cuadratura se aplica tanto a la QPSK como a la QPSK de compensación. Por tanto, podemos afirmar que la probabilidad de error en el canal en fase o en cuadratura de un receptor QPSK de compensación coherente sigue siendo igual a $(1/2)\text{erfc}(\sqrt{E/2N_0})$. En consecuencia, la fórmula de la ecuación (6.34) se aplica igualmente bien a la QPSK de compensación.

■ QPSK DESFASADA π/4

Una señal QPSK ordinaria tal vez resida en cualquiera de las dos constelaciones que suelen usarse y que se muestran en las figuras 6.11*a* y 6.11*b*, las cuales están desfasadas π/4 radianes una de la otra. En otra variante de la QPSK conocida como *QPSK desfasada* π/4,[3] la fase portadora que se usa para la transmisión de símbolos sucesivos (es decir, dibits) se captura alternativamente de una de las dos constelaciones QPSK de la figura 6.11 y luego de la otra. Por tanto, se concluye que una señal QPSK desfasada π/4 reside en cualquiera de ocho posibles estados de fase, como se indica en la



a) *b)*

FIGURA 6.11 Dos constelaciones de señales utilizadas comúnmente para la QPSK; las flechas indican las trayectorias a lo largo de las cuales el modulador QPSK puede cambiar su estado.

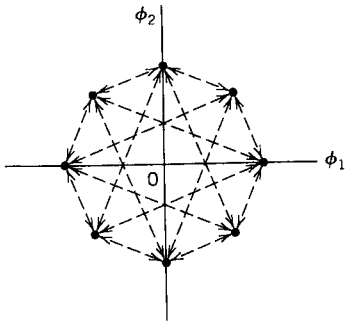


FIGURA 6.12 Ocho estados de fase posibles para la QPSK desfasada $\pi/4$.

figura 6.12. Las cuatro líneas punteadas que surgen de cada punto de mensaje posible en la figura 6.12 definen las transiciones de fase que son factibles en la QPSK desfasada $\pi/4$.

La tabla 6.2 resume un posible conjunto de relaciones entre las transiciones de fase en este nuevo esquema de modulación digital y los dibits de codificación Gray entrantes. Por ejemplo, si el modulador está en uno de los estados de fase indicados en la figura 6.11*b*, entonces en la recepción el dibit 00 se recorre a un estado de fase señalado en la figura 6.11*a* mediante una rotación de $\pi/4$ radianes en una dirección contraria a la de las manecillas del reloj.

Algunas características atractivas del esquema de la QPSK desfasada $\pi/4$ incluyen las siguientes:

- Las transiciones de fase de un símbolo al siguiente están restringidas a $\pm\pi/4$ y $\pm 3\pi/4$ radianes, lo cual contrasta con las transiciones de fase $\pm\pi/2$ y $\pm\pi$ en la QPSK. En consecuencia, las variaciones de envolvente de las señales de la QPSK desfasada $\pi/4$ debidas a filtrado se reducen en forma significativa, en comparación con las correspondientes a la QPSK.
- A diferencia de las señales QPSK de compensación, las correspondientes a la QPSK desfasada $\pi/4$ pueden detectarse de modo no coherente, simplificando por ello de manera considerable el diseño del receptor. Además, al igual que las señales QPSK, las de la QPSK desfasada $\pi/4$ pueden codificarse de manera diferente, en cuyo caso debemos hablar realmente de la *DQPSK desfasada* $\pi/4$.

La generación de símbolos de la DQPSK desfasada $\pi/4$, representados por el par de símbolos (I, Q), se describe mediante el siguiente par de relaciones (vea el problema 6.13):

$$\begin{aligned} I_k &= \cos(\theta_{k-1} + \Delta\theta_k) \\ &= \cos\theta_k \end{aligned} \tag{6.43}$$

$$\begin{aligned} Q_k &= \operatorname{sen}(\theta_{k-1} + \Delta\theta_k) \\ &= \operatorname{sen}\theta_k \end{aligned} \tag{6.44}$$

TABLA 6.2 ***Correspondencia entre el cambio de fase y el dibit de entrada en la DQPSK desfasada $\pi/4$***

Dibit de entrada con codificación Gray	*Cambio de fase, $\Delta\theta$ (radianes)*
00	$\pi/4$
01	$3\pi/4$
11	$-3\pi/4$
10	$-\pi/4$

TABLA 6.3 *Resultados de la DQPSK desfasada π/4 para el ejemplo 6.2*

Escalón k	*Fase* θ_{k-1} *(radianes)*	*Dibit de entrada*	*Cambio de fase* $\Delta\theta_k$ *(radianes)*	*Fase transmitida* θ_k *(radianes)*
1	π/4	00	π/4	π/2
2	π/2	10	−π/4	π/4
3	π/4	10	−π/4	0
4	0	01	3π/4	3π/4

donde θ_{k-1} es el ángulo de fase absoluto del símbolo $k-1$, y $\Delta\theta_k$ es el cambio de fase codificado diferencialmente que se define de acuerdo con la tabla 6.2.

EJEMPLO 6.2

Continuando con la secuencia binaria de entrada del ejemplo 6.1, es decir, 01101000, suponga que el ángulo de fase $\theta_0 = \pi/4$ en la constelación de la figura 6.11*b* se asigna como el estado de fase *inicial* del modulador DQPSK desfasada π/4. En ese caso, al arreglar la secuencia binaria de entrada como una secuencia de dibits y siguiendo la convención de la tabla 6.2, obtenemos los resultados que se presentan en la tabla 6.3 para el ejemplo presente. ◀

Detección de señales DQPSK desfasadas π/4

Luego de habernos familiarizado con la generación de las señales DQPSK desfasadas π/4, consideraremos su detección diferencial. Dada la salida de canal ruidosa $x(t)$, el receptor calcula primero las proyecciones de $x(t)$ sobre las funciones básicas de $\phi_1(t)$ y $\phi_2(t)$. Las salidas que se producen, denotadas por I y Q, respectivamente, se aplican en el *detector diferencial* que consta de los siguientes componentes, como se indica en la figura 6.13:

- *Computadora de arco tangente* para extraer el ángulo de fase θ de la salida del canal (señal recibida).
- *Computadora de diferencia de fase* para determinar el cambio en la fase θ que ocurre sobre un intervalo de símbolo.
- *Lógica de corrección de módulo 2π* para corregir errores debidos a la posibilidad de ángulos de fase que se produzcan alrededor del eje real.

Elaborando un poco más el último punto, sea $\Delta\theta_k$ la diferencia de fase calculada entre θ_k y θ_{k-1} que representa los ángulos de fase de la salida del canal mediante los símbolos de k y $k-1$, respectivamente. Entonces la lógica de corrección del módulo 2π opera como sigue:

$$\begin{aligned} &\text{Si } \Delta\theta_k < -180 \text{ grados Entonces } \Delta\theta_k = \Delta\theta_k + 360 \text{ grados} \\ &\text{Si } \Delta\theta_k > 180 \text{ grados Entonces } \Delta\theta_k = \Delta\theta_k - 360 \text{ grados} \end{aligned} \tag{6.45}$$

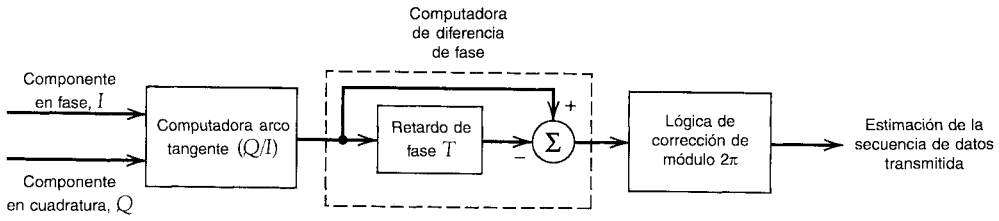


FIGURA 6.13 Diagrama de bloques de la DQPSK desfasada π/4.

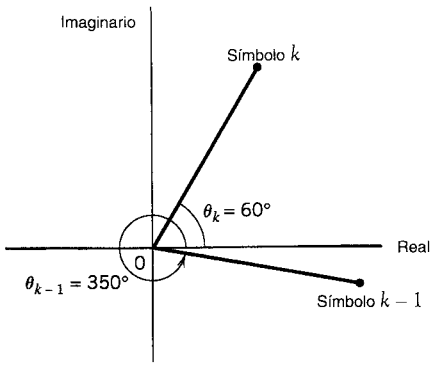


FIGURA 6.14 Ilustración de la posibilidad de ángulos de fase que se envuelven en torno al eje real positivo.

Para ilustrar la necesidad de esta corrección de fase, considere la situación que se presenta en la figura 6.14, donde $\theta_{k-1} = 350$ grados y $\theta_k = 60$ grados, midiéndose ambos ángulos de fase en la dirección contraria a la de las manecillas del reloj. Con base en esta figura vemos fácilmente que el cambio de fase $\Delta\theta_k$, medido en la dirección contraria a la de las manecillas del reloj, es de 70 grados. Sin embargo, sin corrección el ángulo de fase $\Delta\theta_k$ se calcula como 60-350 grados = −290 grados. Aplicando la primera línea de la ecuación (6.45), la lógica de corrección de módulo 2π compensa el envolvimiento alrededor del eje real positivo, produciendo el resultado corregido

$$\Delta\theta_k = -290 \text{ grados} + 360 \text{ grados} = 70 \text{ grados}$$

Resulta relativamente simple la puesta en práctica del detector diferencial de tipo tangente de la figura 6.13 para la demodulación de las señales DQPSK desfasadas $\pi/4$. Este dispositivo ofrece un desempeño satisfactorio en un canal de desvanecimiento de Rayleigh como un ambiente de multitrayectoria estático. Sin embargo, cuando el ambiente de multitrayectoria varía en el tiempo, como sucede en un sistema de comunicación digital e inalámbrico comercial, los resultados de simulación de computadora muestran aparentemente que el desempeño del receptor se degrada muy rápido.[4]

PSK *M*-ARIA

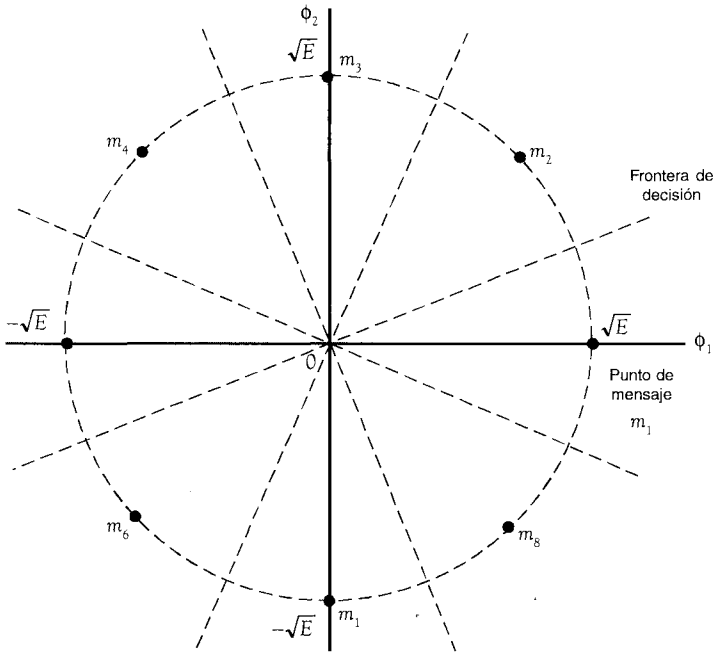
La QPSK es un caso especial de la *PSK M-aria,* donde la fase portadora toma uno de M valores posibles, a saber, $\theta_i = 2(i-1)\pi/M$, donde $i = 1, 2, \ldots, M$. Por tanto, durante cada intervalo de transmisión de señales de duración T, se envía una de las M señales posibles

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + \frac{2\pi}{M}(i-1)\right), \quad i = 1, 2, \ldots, M \tag{6.46}$$

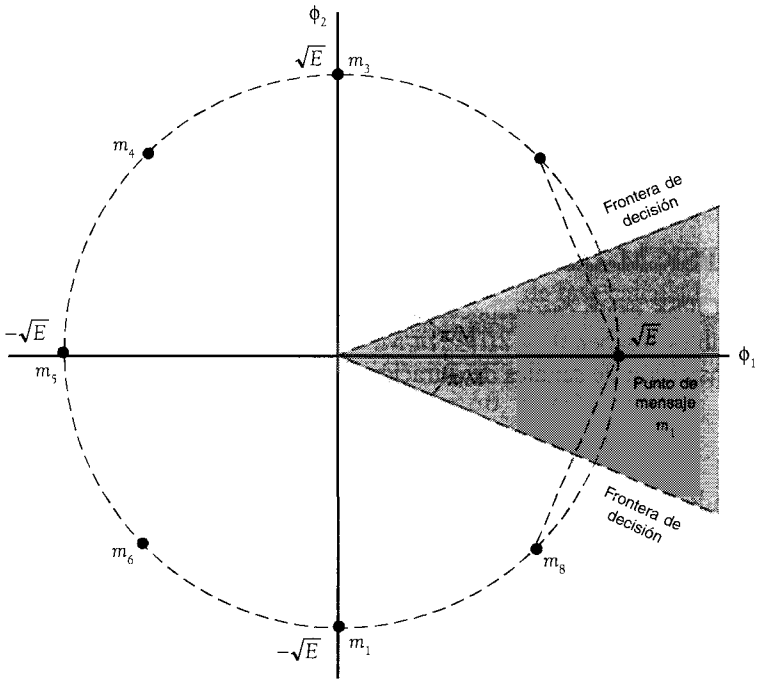
donde E es energía de la señal por símbolo. La frecuencia de la portadora es $f_c = n_c/T$ para algún entero fijo n_c.

Cada $s_i(t)$ puede expandirse en términos de las mismas dos funciones básicas $\phi_1(t)$ y $\phi_2(t)$ definidas en las ecuaciones (6.25) y (6.26), respectivamente. La constelación de señales de la PSK M-aria es, en consecuencia, bidimensional. Los M puntos de mensaje están igualmente espaciados sobre un círculo de radio $\sqrt{E}$ y centro en el origen, como se ilustra en la figura 6.15*a* para el caso de *corrimiento de octofase por llaveo* (es decir, M = 8).

A partir de la figura 6.15*a* advertimos que el diagrama del espacio de señales es circularmente simétrico. Por ello, es posible aplicar la ecuación (5.92), basada en la cota de unión, para desarrollar una fórmula aproximada de la probabilidad promedio del error de símbolo relativa a la PSK M-aria. Suponga que la señal transmitida corresponde al punto de mensaje m_1, cuyas



a)



b)

FIGURA 6.15 *a)* Diagrama del espacio de las señales para el corrimiento de octofase por llaveo (M = 8). Las fronteras de decisión se muestran como líneas discontinuas. *b)* Diagrama del espacio de las señales que ilustra la aplicación de la cota de unión para la codificación por corrimiento de octofase.

coordenadas a lo largo de los ejes ϕ_1 y ϕ_2 son, respectivamente, $+\sqrt{E}$ y 0. Suponga que la proporción E/N_0 es lo suficientemente grande para considerar los dos puntos de mensaje más cercanos, a cualquier lado de m_1, como candidatos potenciales para confundirse con m_1 debido al ruido del canal. Esto se ilustra en la figura 6.15*b* para el caso de M = 8. La distancia euclidiana de cada uno de estos dos puntos a partir de m_1 es (para M = 8)

$$d_{12} = d_{18} = 2\sqrt{E}\,\text{sen}\left(\frac{\pi}{M}\right)$$

Por tanto, el empleo de la ecuación (5.92) del capítulo 5 produce la probabilidad promedio del error de símbolo para la PSK M-aria coherente como

$$P_e \simeq \text{ferc}\left(\sqrt{\frac{E}{N_0}}\,\text{sen}\left(\frac{\pi}{M}\right)\right) \tag{6.47}$$

donde se supone que $M \geq 4$. La aproximación se vuelve en extremo ajustada para M fija, cuando E/N_0 aumenta. Para M = 4 la ecuación (6.47) se reduce a la misma forma dada en la ecuación (6.34) para la QPSK.

Espectro de potencia de las señales PSK M-arias

La duración del símbolo de la PSK M-aria está definida por

$$T = T_b \log_2 M \tag{6.48}$$

donde T_b es la duración de bit. Procediendo de manera similar a la que se describió para una señal de QPSK, podemos demostrar que la densidad espectral de potencia de la banda base de una señal PSK M-aria está dada por

$$\begin{aligned} S_B(f) &= 2E\ \text{senc}^2(Tf) \\ &= 2E_b \log_2 M \text{senc}^2(T_b f \log_2 M) \end{aligned} \tag{6.49}$$

En la figura 6.16 presentamos la densidad espectral de potencia normalizada $S_B(f)/2E_b$ graficada en función de la frecuencia normalizada fT_b para tres valores diferentes de M, a saber, M = 2, 4, 8.

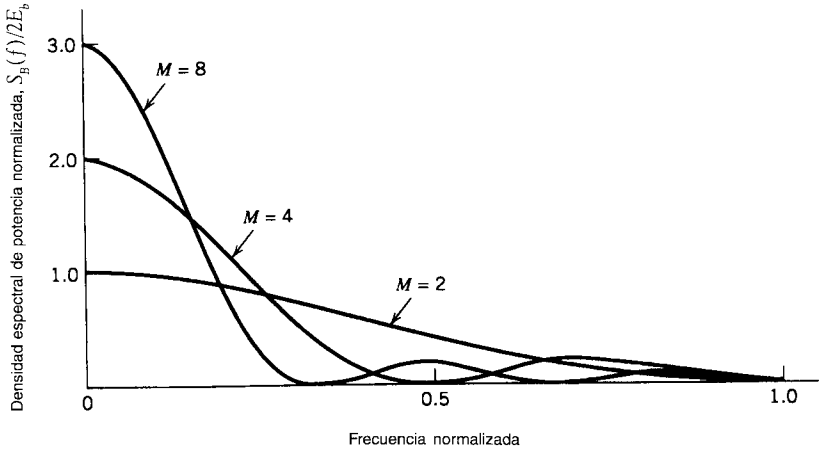


FIGURA 6.16 Espectro de potencia de señales PSK M-aria para M = 2, 4, 8.

TABLA 6.4 *Eficiencia del ancho de banda de las señales PSK M-arias*

M	2	4	8	16	32	64
ρ (bits/s/Hz)	0.5	1	1.5	2	2.5	3

■ EFICIENCIA DEL ANCHO DE BANDA DE SEÑALES PSK M-ARIAS

Los espectros de potencia de las señales PSK M-arias poseen un *lóbulo principal* acotado por *nulos espectrales* (es decir, frecuencias a las cuales la densidad espectral de potencia es cero) perfectamente definidos. De acuerdo con ello, el ancho espectral del lóbulo principal proporciona una medida simple y popular para el ancho de banda de las señales PSK M-arias. Esta definición se conoce como el *ancho de banda de nulo a nulo*. Con el ancho de banda de nulo a nulo abarcando el lóbulo principal del espectro de potencia de una señal M-aria, encontramos que éste contiene la mayor parte de la potencia de la señal. Esto se ve fácilmente al observar las gráficas espectrales de potencia de la figura 3.16.

En las funciones básicas pasobanda que se definen en las ecuaciones (6.25) y (6.26), el ancho de banda del canal que se requiere para dejar pasar señales PSK M-arias (más precisamente, el lóbulo espectral principal de las señales M-arias) está dado por

$$B = \frac{2}{T} \tag{6.50}$$

donde T es la duración del símbolo. Sin embargo, la duración del símbolo T se relaciona con la duración de bit T_b mediante la ecuación (6.48). Asimismo, la tasa de bits es $R_b = 1/T_b$. Por tanto, es posible redefinir el ancho de banda del canal de la ecuación (6.50) en términos de la tasa de bits R_b como

$$B = \frac{2R_b}{\log_2 M} \tag{6.51}$$

Con base en esta fórmula, la eficiencia del ancho de banda de las señales PSK M-arias está dada por

$$\rho = \frac{R_b}{B} = \frac{\log_2 M}{2} \tag{6.52}$$

La tabla 6.4 produce los valores de ρ a partir de la ecuación (6.52) para M variable.

A la luz de la ecuación (6.47) y de la tabla 6.4, podemos establecer la siguiente observación en el contexto de la PSK M-aria: cuando aumenta el número de estados M, se mejora la eficiencia del ancho de banda a costa del desempeño ante los errores. Para asegurar que no haya degradación en el desempeño ante los errores, tenemos que aumentar E_b/N_0 a fin de compensar el aumento en M.

6.4 *Esquemas de modulación híbridos de amplitud/fase*

En un sistema PSK M-ario, las componentes en fase y en cuadratura de la señal modulada se interrelacionan de tal manera que la envolvente está restringida a permanecer constante. Estas restricciones se manifiestan de por sí solas en una constelación circular para los puntos de mensaje. Sin embargo, si se elimina esta restricción y se deja consecuentemente que las compo-

nentes en fase y en cuadratura sean independientes, obtenemos un nuevo esquema de modulación denominado *modulación M-aria* de amplitud en cuadratura (QAM). Este último esquema de modulación es de naturaleza *híbrida,* en el sentido de que la portadora experimenta modulación de amplitud, así como en fase.

Las funciones básicas en pasobanda en la QAM M-aria no pueden ser periódicas para una elección arbitraria de la frecuencia de la portadora f_c con respecto a la velocidad del símbolo $1/T$. Casi en todos los casos, esta aperiodicidad en realidad no preocupa. Al reformular de cierta manera la expresión para la señal transmitida, es posible eliminar la variación del tiempo de las funciones básicas en transmisiones de símbolo sucesivas y, por ello, simplificar la puesta en práctica del transmisor. En particular, se hace que este último aparezca "sin portadora" mientras se mantiene completamente la esencia del proceso híbrido de modulación de amplitud y en fase. Ésta es en realidad la idea detrás de la modulación de amplitud y de fase sin portadora (CAP).

A pesar de las diferencias entre la QAM y la CAP en sus detalles de ejecución, tienen exactamente las mismas constelaciones de señales. Por tanto, resultan fundamentalmente equivalentes en desempeño para una complejidad de receptor preestablecida. En lo que sigue analizamos primero la QAM y después la CAP.

■ Modulación M-aria en amplitud por cuadratura

En los capítulos 4 y 5 estudiamos la modulación M-aria por amplitud de pulsos (PAM), que es adimensional. La QAM M-aria es una generalización bidimensional de la PAM M-aria porque su formulación implica dos funciones básicas pasobanda ortogonales, como se indica mediante

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t), \quad 0 \le t \le T \tag{6.53}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}} \operatorname{sen}(2\pi f_c t), \quad 0 \le t \le T \tag{6.54}$$

Denotemos el punto de mensaje i-ésimo $\mathbf{s}_i$ en el plano (ϕ_1, ϕ_2) mediante $(a_i d_{\text{mín}}/2, b_i d_{\text{mín}}/2)$, donde $d_{\text{mín}}$ es la distancia mínima entre cualesquiera dos puntos de mensaje en la constelación, a_i y b_i son enteros e $i = 1, 2, \ldots, M$. Sea $(d_{\text{mín}}/2) = \sqrt{E_0}$, donde E_0 es la energía de la señal con la amplitud más baja. La señal QAM M-aria que se transmite para el símbolo k, digamos, está definida entonces por

$$s_k(t) = \sqrt{\frac{2E_0}{T}} a_k \cos(2\pi f_c t) - \sqrt{\frac{2E_0}{T}} b_k \operatorname{sen}(2\pi f_c t), \qquad \begin{matrix} 0 \le t \le T \\ k = 0, \pm 1, \pm 2, \ldots \end{matrix} \tag{6.55}$$

La señal $s_k(t)$ consta de dos portadoras en cuadratura de fase que se modulan cada una mediante un conjunto de amplitudes discretas, de ahí el nombre de *modulación en amplitud por cuadratura.*

Dependiendo del número de símbolos posibles M, podemos distinguir dos constelaciones QAM: constelaciones cuadradas para las cuales el número de bits por símbolo es par, y constelaciones cruzadas para las cuales ese número es impar. Estos dos casos se consideran a continuación en el orden mencionado.

Constelaciones cuadradas QAM

Con un número *par* de bits por símbolo, podemos escribir

$$L = \sqrt{M} \tag{6.56}$$

donde L es un entero positivo. Bajo esta condición, una constelación cuadrada QAM M-aria puede considerarse siempre como el *producto cartesiano* de una constelación PAM L-aria unidimensional consigo misma. Por definición, el producto cartesiano de dos conjuntos de co-

ordenadas (representando un par de constelaciones unidimensionales) está conformado por el conjunto de todos los pares ordenados posibles de coordenadas con la primera coordenada de cada uno de ellos tomada del primer conjunto implicado en el producto y la segunda coordenada tomada del segundo conjunto en dicho producto.

En el caso de una constelación cuadrada QAM, los pares ordenados de coordenadas forman naturalmente una matriz cuadrada, como indica

$$\{a_i, b_i\} = \begin{bmatrix} (-L+1, L-1) & (-L+3, L-1) & \dots & (L-1, L-1) \\ (-L+1, L-3) & (-L+3, L-3) & \dots & (L-1, L-3) \\ \vdots & \vdots & & \vdots \\ (-L+1, -L+1) & (-L+3, -L+1) & \dots & (L-1, -L+1) \end{bmatrix} \tag{6.57}$$

EJEMPLO 6.3

Considere una QAM de 16 cuya constelación de señales se describe en la figura 6.17*a*. La codificación de los puntos de mensaje que se muestra en esta figura es la siguiente:

- Dos de los cuatro bits, a saber, los dos más a la izquierda, especifican el cuadrante en el plano (ϕ_1, ϕ_2) en el cual se encuentra el punto de mensaje. De esa manera, empezando desde el primer cuadrante y procediendo en el sentido contrario al de las manecillas del reloj, los cuatro cuadrantes están representados por los dibits 11, 10, 00 y 01.
- Los dos bits restantes se utilizan para representar uno de los cuatro posibles símbolos que se encuentran dentro de cada cuadrante del plano (ϕ_1, ϕ_2)

Advierta que la codificación de los cuatro cuadrantes y la correspondiente también a los símbolos en cada cuadrante sigue la regla de codificación de Gray.

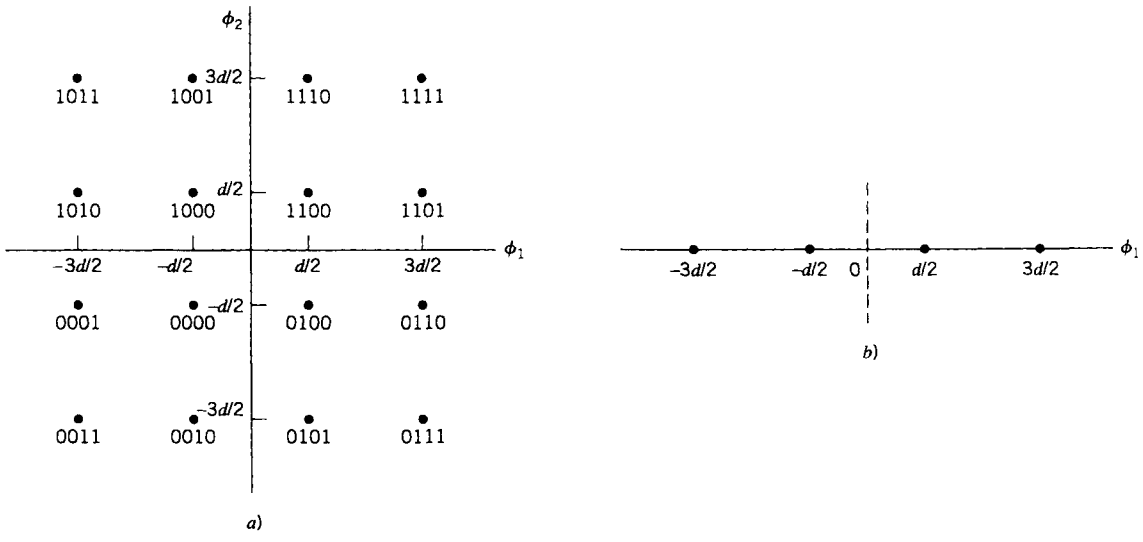


FIGURA 6.17 *a*) Diagramas del espacio de señales de la QAM M-aria para M = 16; los puntos de mensaje en cada cuadrante se identifican con los cuadribits de codificación Gray. *b*) Diagrama del espacio de señales de la señal PAM 4 correspondiente.

En el ejemplo presente tenemos $L = 4$. De ese modo, la constelación cuadrada de la figura 6.17*a* es el producto cartesiano consigo misma de la constelación PAM de 4 que se muestra en la figura 6.17*b*. Asimismo, la matriz de la ecuación (6.57) tiene el valor

$$\{a_i, b_i\} = \begin{bmatrix} (-3,3) & (-1,3) & (1,3) & (3,3) \\ (-3,1) & (-1,1) & (1,1) & (3,1) \\ (-3,-1) & (-1,-1) & (1,-1) & (3,-1) \\ (-3,-3) & (-1,-3) & (1,-3) & (3,-3) \end{bmatrix}$$

Para calcular la probabilidad de error del símbolo para la QAM M-aria explotamos la propiedad de que la constelación cuadrada de la QAM puede factorizarse en el producto de las constelaciones PAM correspondientes consigo mismas. En consecuencia, podemos proceder del modo siguiente:

1. La probabilidad de detección correcta de la QAM M-aria puede escribirse como

$$P_c = (1 - P_e')^2 \tag{6.58}$$

donde P_e' es la probabilidad del error de símbolo para la PAM L-aria correspondiente con $L = \sqrt{M}$.

2. La probabilidad de error del símbolo P_e' está definida por

$$P_e' = \left(1 - \frac{1}{\sqrt{M}}\right) \operatorname{ferc}\left(\sqrt{\frac{E_0}{N_0}}\right) \tag{6.59}$$

(Advierta que $L = \sqrt{M}$ en la QAM M-aria corresponde a la M en la PAM M-aria considerada en el problema 4.27.)

3. La probabilidad del error de símbolo para la QAM M-aria está dada por

$$\begin{aligned} P_e &= 1 - P_c \\ &= 1 - (1 - P_e')^2 \\ &\simeq 2P_e' \end{aligned} \tag{6.60}$$

donde se supone que P_e' es suficientemente pequeña en comparación con la unidad para justificar el que se ignore el término cuadrático.

De ese modo, utilizando las ecuaciones (6.58) y (6.59) en la (6.60), encontramos que la probabilidad de error del símbolo en la QAM M-aria está dada aproximadamente por

$$P_e \simeq 2\left(1 - \frac{1}{\sqrt{M}}\right) \operatorname{ferc}\left(\sqrt{\frac{E_0}{N_0}}\right) \tag{6.61}$$

La energía transmitida en la QAM M-aria es variable en cuanto que su valor instantáneo depende del símbolo particular que se transmite. Por tanto, resulta más lógico expresar P_e en términos del valor *promedio* de la energía transmitida y no en los de E_0. Suponiendo que los L niveles de amplitud de la componente en fase o en cuadratura son igualmente probables, tenemos

$$E_{\text{pro}} = 2\left[\frac{2E_0}{L}\sum_{i=1}^{L/2}(2i-1)^2\right] \tag{6.62}$$

donde el factor multiplicador 2 fuera de los corchetes toma en cuenta las contribuciones iguales hechas de las componentes en fase y en cuadratura. Los límites de la sumatoria y el factor multiplicador 2 dentro de los corchetes explica la naturaleza simétrica de los niveles de amplitud pertinentes alrededor del cero. Al efectuar la sumatoria en la ecuación (6.62) obtenemos

$$\begin{aligned} E_{\text{pro}} &= \frac{2(L^2-1)E_0}{3} \\ &= \frac{2(M-1)E_0}{3} \end{aligned} \tag{6.63}$$

En consecuencia, podemos reescribir la ecuación (6.61) en términos de E_{pro} como

$$P_e \simeq 2\left(1-\frac{1}{\sqrt{M}}\right)\text{ferc}\left(\sqrt{\frac{3E_{\text{pro}}}{2(M-1)N_0}}\right) \tag{6.64}$$

que es el resultado deseado.

El caso de M = 4 es de interés especial. La constelación de señales para este valor de M es la misma que la correspondiente a la QPSK. En realidad, haciendo M = 4 en la ecuación (6.64) y advirtiendo que para este caso especial $E_{\text{pro}} = E$, donde E es la energía por símbolo, encontramos que la fórmula resultante para la probabilidad del error de símbolo se vuelve idéntica a la de la ecuación (6.34), y así debe ser.

Constelación cruzada QAM

Para generar una señal QAM M-aria con un número impar de bits por símbolo, requerimos el uso de una *constelación cruzada*. Como se ilustra en la figura 6.18, es posible construir una constelación de señales de este tipo con n bits por símbolo procediendo del siguiente modo:

- Empezar con una constelación cuadrada QAM con $n - 1$ bits por símbolo.
- Extender cada lado de la constelación cuadrada QAM agregando 2^{n-3} símbolos.
- Ignorar las esquinas en extensión.

El cuadrado interno representa 2^{n-1} símbolos. Las cuatro extensiones de los lados agregan $4 \times 2^{n-3} = 2^{n-1}$ símbolos. El número total de símbolos en la constelación cruzada es entonces $2^{n-1} + 2^{n-1}$, lo cual es igual a 2^n y, por tanto, representa n bits por símbolo como se desea.

A diferencia de la constelación cuadrada QAM, *no* es posible expresar una constelación cruzada QAM como el producto de una constelación PAM consigo misma. La falta de una

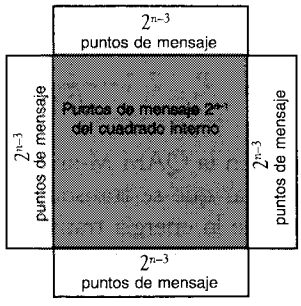


FIGURA 6.18 Ilustración de cómo una constelación QAM al cuadrado puede expandirse para formar una constelación cruzada QAM.

factorización de este tipo complica la determinación de la probabilidad del error de símbolo P_e en el que se incurre al usar la QAM M-aria caracterizada por una constelación cruzada. Por tanto, simplemente establecemos la fórmula para P_e sin demostración, en la forma

$$P_e \simeq 2\left(1-\frac{1}{\sqrt{2M}}\right)\text{ferc}\left(\sqrt{\frac{E_0}{N_0}}\right)\text{para } E_0/N_0 \text{ elevada} \tag{6.65}$$

la cual concuerda con la fórmula de la ecuación (6.61) para una constelación cuadrada, excepto por la inclusión de un 0.5 de bit adicional por dimensión en la constelación.[5] Advierta también que no es posible aplicar el código Gray perfectamente a una constelación cruzada QAM.

Modulación de amplitud/fase sin portadora

Las funciones básicas pasobanda de las ecuaciones (6.53) y (6.54) suponen el uso de un pulso rectangular para la función de formación de pulso. Por razones que serán claras, redefiniremos la señal QAM M-aria transmitida de la ecuación (6.55) en términos de una función general de pulsos $g(t)$ como

$$s_k(t) = a_k g(t-kT)\cos(2\pi f_c t) - b_k g(t-kT)\,\text{sen}(2\pi f_c t), \quad \begin{matrix} 0 \le t \le T \\ k = 0, \pm 1, \pm 2, \ldots \end{matrix} \tag{6.66}$$

Se supone que la frecuencia f_c de la portadora tiene un valor arbitrario con respecto a la velocidad de símbolo $1/T$. Con base en la ecuación (6.66), es posible expresar la señal QAM M-aria transmitida $s(t)$ para una sucesión infinita de símbolos como

$$\begin{aligned} s(t) &= \sum_{k=-\infty}^{\infty} s_k(t) \\ &= \sum_{k=-\infty}^{\infty} \left[a_k g(t-kT)\cos(2\pi f_c t) - b_k g(t-kT)\,\text{sen}(2\pi f_c t)\right] \end{aligned} \tag{6.67}$$

Esta ecuación muestra que para una f_c arbitraria, las funciones pasobanda $g(t-kT)\cos(2\pi f_c t)$ y $g(t-kT)\,\text{sen}(2\pi f_c t)$ son *aperiódicas* con respecto a que varían de un símbolo a otro.

¿Cómo podemos eliminar las variaciones en el tiempo de estas funciones básicas pasobanda de símbolo a símbolo? Para responder a la pregunta, encontramos conveniente cambiar nuestro formalismo de la notación real a la compleja. De modo específico, reescribimos la ecuación (6.67) en la forma equivalente

$$\begin{aligned} s(t) &= \text{Re}\left\{\sum_{k=-\infty}^{\infty} (a_k + jb_k) g(t-kT)\exp(j2\pi f_c t)\right\} \\ &= \text{Re}\left\{\sum_{k=-\infty}^{\infty} A_k g(t-kT)\exp(j2\pi f_c t)\right\} \end{aligned} \tag{6.68}$$

donde A_k es un número complejo definido por

$$A_k = a_k + jb_k \tag{6.69}$$

y $\text{Re}\{\cdot\}$ denota la parte real de la cantidad compleja encerrada dentro de las llaves. Claramente, la ecuación (6.68) permanece sin cambio al multiplicar el sumando de esta ecuación por la unidad

expresada como el producto de la exponencial compleja $\exp(-j2\pi f_c kT)$ y su conjugada compleja $\exp(j2\pi f_c kT)$. De este modo es posible reescribir la ecuación (6.68) en una forma nueva

$$\begin{aligned} s(t) &= \mathrm{Re}\left\{\sum_{k=-\infty}^{\infty} A_k g(t-kT)\exp(j2\pi f_c t)\exp(-j2\pi f_c kT)\exp(j2\pi f_c kT)\right\} \\ &= \mathrm{Re}\left\{\sum_{k=-\infty}^{\infty} A_k \exp(j2\pi f_c kT) g(t-kT)\exp(j2\pi f_c (t-kT))\right\} \end{aligned} \tag{6.70}$$

Defina

$$\tilde{A}_k = A_k \exp(j2\pi f_c kT) \tag{6.71}$$

$$g_+(t) = g(t)\exp(j2\pi f_c t) \tag{6.72}$$

El escalar $\tilde{A}_k$ es simplemente una versión *rotada* de la representación compleja de las coordenadas del símbolo transmitido k-ésimo en el plano (ϕ_1, ϕ_2). Antes de presentar una interpretación de la señal de valores complejos $g_+(t)$, supondremos que la función de formación de pulso $g(t)$ es una señal pasobajas cuya componente de frecuencia más alta es más pequeña que la frecuencia de la portadora f_c. Siguiendo ahora el material que se presenta en el apéndice 2, reconocemos a $g_+(t)$ como la representación de la *señal analítica* o *preenvolvente* de la señal pasobanda $g(t)\cos(2\pi f_c t)$. Para ser más específicos, expandimos $g_+(t)$ en la forma

$$\begin{aligned} g_+(t) &= g(t)\cos(2\pi f_c t) + jg(t)\,\mathrm{sen}(2\pi f_c t) \\ &= p(t) + j\hat{p}(t) \end{aligned} \tag{6.73}$$

donde $p(t)$ y $\hat{p}(t)$ se definen mediante

$$p(t) = g(t)\cos(2\pi f_c t) \tag{6.74}$$

y

$$\hat{p}(t) = g(t)\,\mathrm{sen}(2\pi f_c t) \tag{6.75}$$

Podemos entonces afirmar que la componente en cuadratura (imaginaria) $\hat{p}(t)$ de la señal analítica $g_+(t)$ es la *transformada de Hilbert* de la componente en fase (real) $p(t)$. Advierta que mientras que la función de formación de pulsos $g(t)$ es una función de la banda base, las componentes en fase y en cuadratura de la señal analítica correspondiente $g_+(t)$ son ambas funciones pasobanda. De aquí en adelante, $g(t)$se referirá como el *pulso de banda base,* y $p(t)$ y $\hat{p}(t)$, como los pulsos *en fase pasobanda* y *en cuadratura pasobanda,* respectivamente.

Con las definiciones de $\tilde{A}_k$ y $g_+(t)$ presentes, estamos listos para redefinir finalmente la señal transmitida de la ecuación (6.70) simplemente como

$$s(t) = \mathrm{Re}\left\{\sum_{k=-\infty}^{\infty} \tilde{A}_k g_+(t-kT)\right\} \tag{6.76}$$

Vale la pena hacer tres importantes observaciones a partir de esta nueva formulación de la señal transmitida:

- La señal transmitida $s(t)$ aparece sin portadora.
- Puesto que la formulación de $s(t)$ en la ecuación (6.76) es en realidad equivalente a la de la señal QAM M-aria presentada en la ecuación (6.67), la nueva formulación de $s(t)$ retiene por completo la modulación híbrida de amplitud y de fase que caracteriza a la señal QAM M-aria original.
- La señal transmitida $s(t)$ representa una realización del símbolo invariante con el tiempo de este proceso de modulación híbrido.

Para una frecuencia de portadora f_c preestablecida, se conoce la secuencia de rotaciones descritas en la ecuación (6.71). Por tanto, el receptor sólo necesita detectar $\widetilde{A}_k$, en cuyo caso podemos calcular el valor de A_k aplicando las rotaciones inversas como se describe a continuación:

$$A_k = \widetilde{A}_k \exp(-j2\pi f_c kT), \quad k = 0, \pm 1, \pm 2, \ldots$$

Sin embargo, en la práctica se ignoran las rotaciones, ya que éstas *no* tienen ninguna influencia en la operación o el desempeño del sistema de modulación híbrido; la aplicación de las rotaciones es en realidad necesaria sólo cuando su equivalencia con la QAM es un asunto de interés. Por tanto, podemos ignorar la ecuación (6.71) y redefinir la señal transmitida de la ecuación (6.76) simplemente como

$$\begin{aligned} s(t) &= \mathrm{Re}\left\{ \sum_{k=-\infty}^{\infty} A_k g_+(t - kT) \right\} \\ &= \mathrm{Re}\left\{ \sum_{k=-\infty}^{\infty} (a_k + jb_k)(p(t - kT) + j\hat{p}(t - kT)) \right\} \\ &= \sum_{k=-\infty}^{\infty} \left[a_k p(t - kT) - b_k \hat{p}(t - kT) \right] \end{aligned} \tag{6.77}$$

donde, como se mencionó anteriormente, $\hat{p}(t)$ es la transformada de Hilbert de $p(t)$. Por razones obvias, la señal transmitida de la ecuación (6.77) se conoce como *modulación de amplitud/fase sin portadora* (CAP).[6]

Propiedades de los pulsos pasobanda en fase y en cuadratura

A partir de las definiciones de los pulsos pasobanda en fase y en cuadratura dadas en las ecuaciones (6.74) y (6.75), deducimos las siguientes propiedades:

Propiedad 1

El pulso pasobanda en fase p(t) *y el pulso en cuadratura* $\hat{p}(t)$ *son funciones par e impar del tiempo* t, *respectivamente, puesto que el pulso de banda base* g(t) *es una función par del tiempo* t.

Esta propiedad se desprende directamente de las ecuaciones (6.74) y (6.75).

Propiedad 2

Los pulsos pasobanda p(t) y $\hat{p}(t)$ *forman un conjunto ortogonal sobre el intervalo completo* $(-\infty, \infty)$ *como indica*

$$\int_{-\infty}^{\infty} p(t)\hat{p}(t)\, dt = 0 \tag{6.78}$$

La razón para no restringir el intervalo de integración en la ecuación (6.78) a un periodo del símbolo T es que los pulsos $p(t)$ y $\hat{p}(t)$ son eficientes en el ancho de banda en sistemas CAP de alto desempeño. Para demostrar la propiedad 2, transformamos la ecuación (6.78) al dominio de la frecuencia utilizando la transformada de Fourier para escribir

$$\int_{-\infty}^{\infty} p(t)\hat{p}(t)\, dt = \int_{-\infty}^{\infty} P(f)\hat{P}^*(f)\, df \tag{6.79}$$

donde $p(t) \rightleftharpoons P(f)$ y $\hat{p}(t) \rightleftharpoons \hat{P}(f)$. El asterisco en $\hat{P}^*(f)$ denota conjugación compleja. (La función de frecuencia $\hat{P}(f)$ *no* debe considerarse como la transformada de Hilbert de $P(f)$; más

bien, siguiendo la terminología usual en el análisis de Fourier, $\hat{P}(f)$ es simplemente la transformada de Fourier de $\hat{p}(t)$. La transformada de Fourier $\hat{P}(f)$ se relaciona con la transformada de Fourier $P(f)$ mediante (vea el apéndice 2)

$$\hat{P}(f) = -j\,\mathrm{sgn}(f)P(f) \tag{6.80}$$

donde $\mathrm{sgn}(f)$ es la función signo. Las transformadas de Fourier $P(f)$ y $\hat{P}(f)$ tienen el mismo espectro de magnitud, aunque sus espectros de fase difieren en +90 grados para frecuencias negativas y en –90 grados para frecuencias positivas. Por tanto, podemos reescribir la ecuación (6.79) como

$$\begin{aligned}\int_{-\infty}^{\infty} p(t)\hat{p}(t)\,dt &= j\int_{-\infty}^{\infty} P(f)P^*(f)\,\mathrm{sgn}(f)\,df \\ &= j\int_{-\infty}^{\infty} |P(f)|^2\,\mathrm{sgn}(f)\,df\end{aligned} \tag{6.81}$$

Admitiendo que la respuesta en magnitud $|P(f)|$ es una función par de la frecuencia f y la función signo $\mathrm{sgn}(f)$ es una función impar de la frecuencia f, la integral de la ecuación (6.81) es cero, lo que demuestra la propiedad 2.

A continuación dejemos que los pulsos pasobanda $p(t)$ y $\hat{p}(t)$ pasen por un canal invariante con el tiempo de respuesta al impulso $h(t)$, produciendo las siguientes salidas pasobanda:

$$u(t) = p(t) \bigstar h(t) \tag{6.82}$$

y

$$\hat{u}(t) = \hat{p}(t) \bigstar h(t) \tag{6.83}$$

donde el símbolo ★ denota convolución. Podemos formular entonces la tercera propiedad.

Propiedad 3

Los pulsos pasobanda u(t) *y* û(t), *definidos en las ecuaciones (6.82) y (6.83), forman un par de transformadas de Hilbert y son, por tanto, ortogonales sobre el intervalo completo* $(-\infty, \infty)$ *para toda* h(t).

Esta importante propiedad posibilita que el receptor CAP separe los símbolos transmitidos real e imaginario, a_k y b_k, dada la salida del canal. Para demostrar la propiedad 3 recurrimos de nuevo a la transformada de Fourier [de manera similar a las ecuaciones (6.79) y (6.81)] para escribir

$$\begin{aligned}\int_{-\infty}^{\infty} u(t)\hat{u}(t)\,dt &= \int_{-\infty}^{\infty} U(f)\hat{U}^*(f)\,df \\ &= \int_{-\infty}^{\infty} (P(f)H(f))(-j\,\mathrm{sgn}(f)P(f)H(f))^*\,df \\ &= \int_{-\infty}^{\infty} j|P(f)|^2|H(f)|^2\,\mathrm{sgn}(f)\,df \\ &= 0\end{aligned}$$

▶ EJEMPLO 6.4 Formación espectral de ancho de banda eficiente

La señal CAP de la ecuación (6.77) utiliza codificación de multinivel por medio del escalar complejo A_k para la eficiencia espectral. Es posible mejorar aún más la eficiencia del ancho de banda de la CAP utilizando una formulación espectralmente eficiente del pulso de banda base $g(t)$. Para la selección de $g(t)$ podemos recurrir a la *familia del coseno elevado* de las funciones de formación de pulsos que se mencionaron en el capítulo 4.

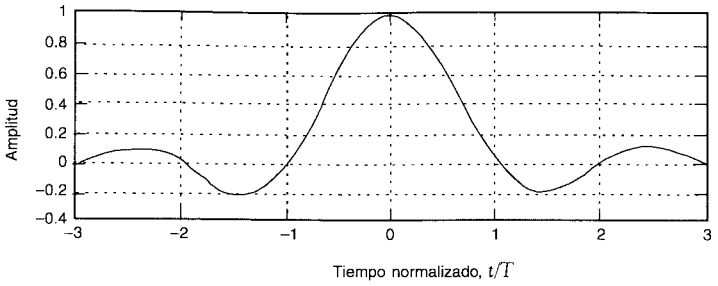


FIGURA 6.19 El pulso de banda base $g(t)$ para el factor de caída $\alpha = 0.2$.

Considere, por ejemplo, un filtro de formación de coseno elevado de banda base con un factor de pendiente de caída $\alpha = 0.2$ (es decir, ancho de banda de exceso de 20%). La fórmula para el pulso de banda base $g(t)$ es*

$$g(t) = (\text{senc}(t))\frac{\cos(\pi\alpha t)}{1-4\alpha^2 t^2} \tag{6.84}$$

donde el tiempo t está normalizado con respecto a la duración del símbolo T. Por tanto, dejando $\alpha = 0.2$ en la ecuación (6.84), obtenemos la gráfica de la figura 6.19. Utilizando las ecuaciones (6.74) y (6.75), es posible calcular el pulso en fase $p(t)$ y el pulso en cuadratura $\hat{p}(t)$ pasobanda graficados en las figuras 6.20*a* y 6.20*b* para una frecuencia de portadora normalizada $f_c T = 0.5(1 + \alpha) = 0.6$. Las formas de onda de la figura 6.20 muestran que $p(t)$ y $\hat{p}(t)$ son funciones par e impar del tiempo t, respectivamente, de acuerdo con la propiedad 1 y son ortogonales sobre el intervalo $(-\infty, \infty)$, lo que concuerda con la propiedad 2.

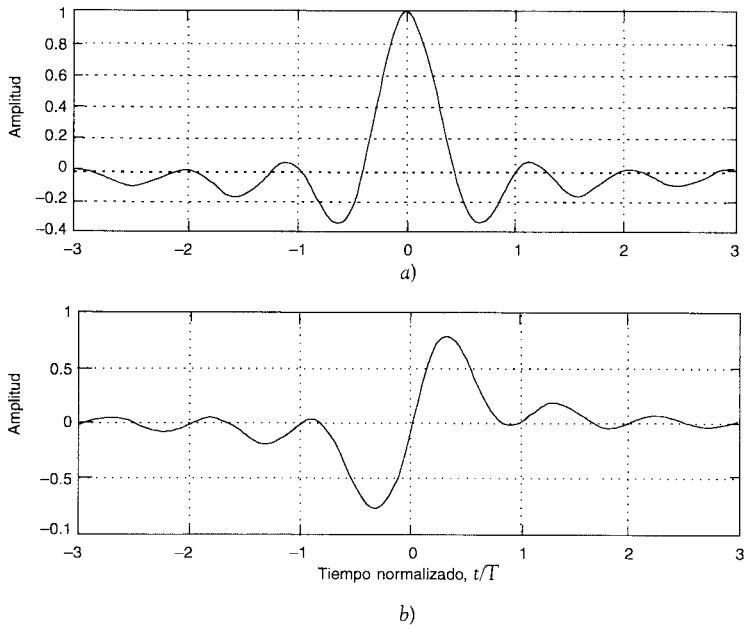


FIGURA 6.20 *a*) Pulso en fase $p(t)$, y *b*) pulso en cuadratura $\hat{p}(t)$ para el factor de rodada $\alpha = 0.2$.

*La ecuación (6.84) se obtiene utilizando el tiempo normalizado t en lugar de $2Wt$ en la ecuación (4.62) del capítulo 4. Recuerde que el ancho de banda de Nyquist W es igual a 0.5 veces la tasa (bits) del símbolo. Advierta que también la $p(t)$ del capítulo 4 no debe confundirse con la $p(t)$ de este capítulo.

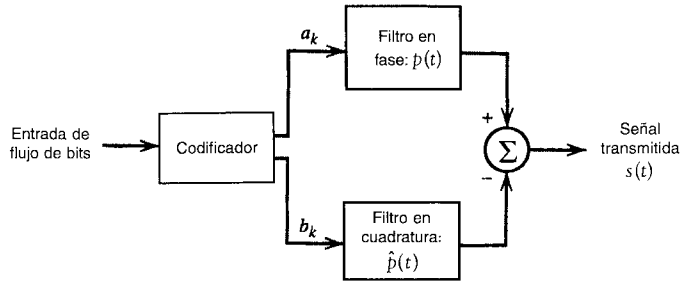


FIGURA 6.21 Diagrama de bloques de un transmisor CAP.

Estructura básica del sistema CAP

La figura 6.21 muestra un diagrama de bloques funcionales del transmisor CAP, el cual se integra con base en el material que abarcan las ecuaciones (6.74), (6.75) y (6.77). El transmisor está compuesto por un codificador multinivel y por un par de filtros pasobanda. El codificador *multinivel* divide la secuencia de datos en serie entrante en bloques sucesivos cada uno de n bits; estos bloques, a su vez, se hacen corresponder con los símbolos multinivel a_k y b_k, donde k se refiere al periodo del símbolo k-ésimo. Los *filtros pasobanda en fase* y *en cuadratura* procesan en paralelo las secuencias de símbolos $\{a_k\}$ y $\{b_k\}$, respectivamente. Las respuestas al impulso de estos dos filtros, a saber, $p(t)$ y $\hat{p}(t)$, se diseñan de acuerdo con las ecuaciones (6.74) y (6.75), para un pulso $g(t)$ de banda base elegido. Las salidas que producen estos dos filtros se restan para producir la señal CAP transmitida $s(t)$ en concordancia con la ecuación (6.77).

La señal transmitida $s(t)$ se propaga por un canal caracterizado por respuesta al impulso $h(t)$ y ruido aditivo $w(t)$. La salida de canal que resulta está definida por

$$\begin{aligned} x(t) &= s(t) \bigstar h(t) + w(t) \\ &= \sum_{k=-\infty}^{\infty} \left[a_k p(t-kT) \bigstar h(t) - b_k \hat{p}(t-kT) \bigstar h(t)\right] + w(t) \\ &= \sum_{k=-\infty}^{\infty} \left[a_k u(t-kT) - b_k \hat{u}(t-kT)\right] + w(t) \end{aligned} \tag{6.85}$$

donde $u(t)$ y $\hat{u}(t)$ están determinadas por las ecuaciones (6.82) y (6.83), respectivamente. Dada $x(t)$ como la señal de entrada, la función del receptor radica en recuperar los símbolos transmitidos a_k y b_k de un modo óptimo en un esquema de símbolo por símbolo.

El receptor consiste en un receptor óptimo bidimensional, cuyos detalles dependen de los deterioros del canal. En forma específica, es posible mencionar dos situaciones diferentes:

- El *ruido blanco gaussiano aditivo es el único deterioro.* En esta situación idealizada, el receptor CAP óptimo está integrado por un *filtro de acoplamiento bidimensional,* como se indica

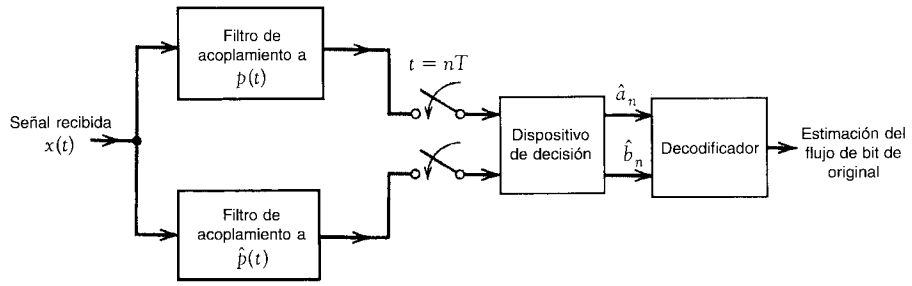


FIGURA 6.22 Diagrama de bloques del receptor CAP utilizando un filtro de acoplamiento bidimensional para el caso de un canal con ruido blanco gaussiano.

en la figura 6.22. El filtro en fase óptimo consiste en un filtro de acoplamiento al pulso en fase pasobanda $p(t)$. El filtro en cuadratura óptimo consiste en un filtro de acoplamiento al pulso en cuadratura pasobanda $\hat{p}(t)$.

- *La interferencia de intersímbolo y el ruido blanco gaussiano aditivo son los deterioros del canal.* En esta situación más realista en términos prácticos, la estructura del receptor CAP óptimo se desprende de la teoría del receptor lineal óptimo que se presentó en la sección 4.9. En particular, el receptor CAP se compone de un filtro de acoplamiento bidimensional, seguido por un par de ecualizadores idénticos y por un par de muestreadores síncronos, como se muestra en la figura 6.23. Los ecualizadores (puestos en práctica en la forma de línea de retardo en derivación) compensan la dispersión en el canal.

De acuerdo con la propiedad 3, deducimos que la salida de un filtro de acoplamiento en fase debida al pulso en cuadratura pasobanda es cero y viceversa. Por tanto, en ambas situaciones, los símbolos transmitidos a_k y b_k pueden detectarse por separado mediante el receptor CAP óptimo bidimensional.

Implementación digital del receptor CAP

Es posible simplificar aún más la estructura del receptor de la figura 6.23 reformulándolo primero en la forma que se muestra en la figura 6.24*a*, donde hemos introducido un convertidor analógico-digital (A/D) a la entrada del receptor para facilitar el uso de un procesamiento de señales digital. A continuación admitimos que los filtros de acoplamiento y los ecualizadores en la figura 6.24*a* son en su totalidad sistemas lineales. En consecuencia, las muestras en los puntos A y B en la figura 6.24*a* son *combinaciones lineales* de las muestras de entrada $x[n] = x(nT_s)$, donde T_s es el periodo de muestreo. Por tanto, se concluye que es posible sustituir la combinación de filtro de acoplamiento y ecualizador en cada trayectoria del receptor en la figura 6.24*a* por un único *filtro de respuesta al impulso de duración finita* (FIR) operando a la frecuencia de muestreo $1/T_s$, como se muestra en la figura 6.24*b*. Con una entrada común, es natural para los dos filtros FIR de la figura 6.24*b* compartir un conjunto de elementos de retardo unitario, pero diferentes conjuntos de coeficientes propios. Advierta también que la estructura del receptor de la figura 6.24*b* es la imagen de espejo de la estructura de transmisor de la figura 6.21.

Conviene un último comentario. En la práctica, los dos filtros FIR en el receptor de la figura 6.24*b* se hacen *adaptables*, de modo que acepten la operación del receptor CAP en un ambiente desconocido. Con la ecuación (filtrado) adaptable como el método elegido, tenemos dos opciones de acuerdo con el material presentado en el capítulo 4: ecualización lineal y ecualización de decisión realimentada (DFE). Cuando la respuesta en frecuencia del canal es aproximadamente plana, el uso de ecualización lineal resulta adecuado para la tarea presente. Sin embargo, cuando la respuesta en frecuencia del canal y/o el espectro de potencia del ruido no son aproximadamente planos, el receptor CAP puede mejorarse de manera considerable mediante el uso de la ecualización de decisión realimentada.

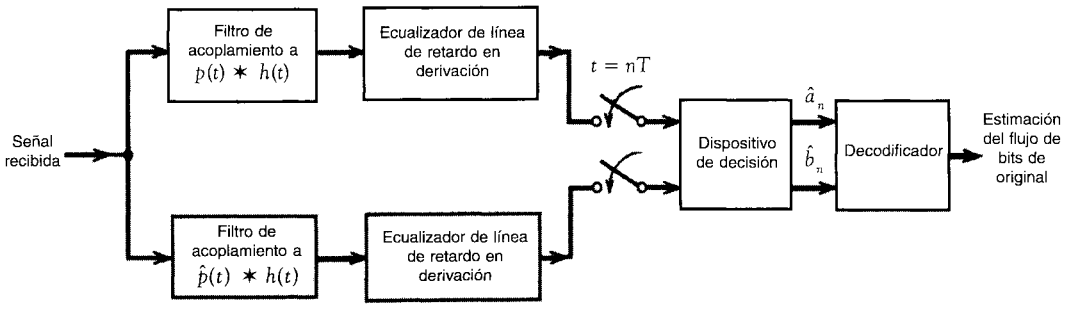


FIGURA 6.23 Diagrama de bloques del receptor CAP utilizando un par de receptores lineales óptimos para el caso de un canal disperso ruidoso.

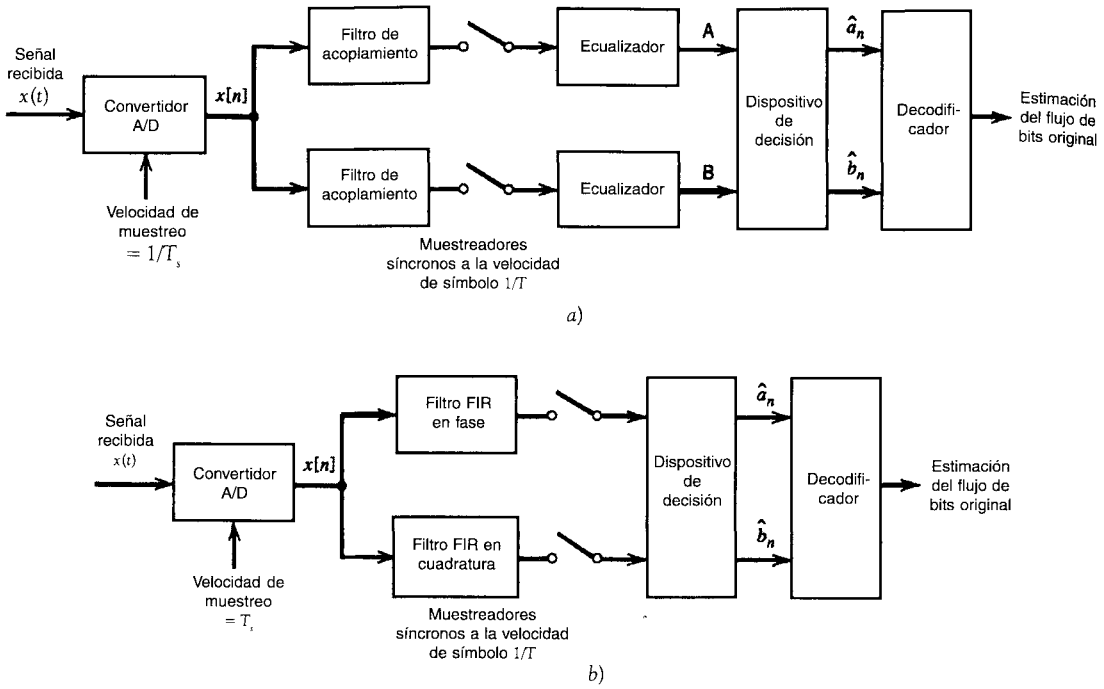


FIGURA 6.24 Puesta en práctica digital del receptor CAP. *a*) Reformulación del receptor en la figura 6.23 utilizando un convertidor A/D y la circuitería asociada. *b*) Sustitución de los pares filtro/ecualizador de acoplamiento con filtros FIR equivalentes (puestos en práctica en forma digital).

Una aplicación de la CAP

Una aplicación importante de la CAP (utilizando 32 o 64 puntos de constelación) se encuentra en la transmisión pasobanda de datos digitales por alambrado de par trenzado de longitud menor que 100 m, como en las redes de área local (LAN) adecuadas para sistemas de distribución de equipos e instalaciones. (Todos los estándares de LAN modernos limitan la longitud del alambrado a un máximo de 100 m; los sistemas CAP tienen la posibilidad de operar realmente en lazos más largos, si así se requiere.) La velocidad de datos puede variar de 51 hasta 155 Mb/s, con el ancho de banda del canal útil limitándose estrictamente a 30 MHz.

Los dos principales deterioros para los transceptores que proporcionan operación dúplex por cables de par trenzado son la pérdida de propagación y la diafonía de extremos cercanos (NEXT); estos deterioros se explicaron en la sección 4.8, la cual trató las líneas digitales de suscriptor.

En el ambiente descrito aquí, un aspecto práctico corresponde a cómo adaptar el receptor a variaciones amplias en los pares de par trenzado. Para satisfacer este requerimiento se recurre a los filtros adaptables para poner en práctica los filtros óptimos en fase y en cuadratura en el receptor, como se subrayó antes. Cuanto más ancho sea el intervalo de cables de par trenzado que pueda alojarse, tanto más complejos deben ser estos filtros adaptativos.

6.5 *Corrimiento de frecuencia por llaveo coherente*

La PSK M-aria y la QAM M-aria comparten una propiedad común: ambas constituyen ejemplos de modulación lineal. En esta sección estudiamos un método no lineal de transmisión de datos pasobanda, a saber, el *corrimiento de frecuencia por llaveo* (FSK) *coherente*. Empezamos el estudio considerando el caso simple de la FSK binaria.

FSK BINARIA

En un *sistema FSK binario*, los símbolos 1 y 0 se distinguen uno del otro transmitiendo una de dos ondas senoidales que difieren en frecuencia por una cantidad fija. Un par característico de ondas senoidales se describe por medio de

$$s_i(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}} \cos(2\pi f_i t), & 0 \le t \le T_b \\ 0, & \text{en otro lado} \end{cases} \tag{6.86}$$

donde $i = 1, 2$, y E_b es la energía por bit de la señal transmitida; la frecuencia transmitida es

$$f_i = \frac{n_c + i}{T_b} \quad \text{para algún entero fijo } n_c \text{ e } i = 1, 2 \tag{6.87}$$

De ese modo, el símbolo 1 se representa por medio de $s_1(t)$ y el símbolo 0 por medio de $s_2(t)$. La señal FSK que se describe aquí se conoce como *FSK de Sunde*. Ésta es una *señal de fase continua* en el sentido de que la continuidad de fase siempre se mantiene, incluyendo los tiempos de conmutación de interbit. Esta forma de modulación digital constituye un ejemplo de *corrimiento de frecuencia por llaveo en fase continua* (CPFSK), sobre la cual después se comentará más en esta sección.

De acuerdo con las ecuaciones (6.86) y (6.87), observamos directamente que las señales $s_1(t)$ y $s_2(t)$ son ortogonales, pero no están normalizadas para tener energía unitaria. Por tanto, deducimos que la forma más útil para el conjunto de funciones básicas ortogonales es

$$\phi_i(t) = \begin{cases} \sqrt{\dfrac{2}{T_b}} \cos(2\pi f_i t), & 0 \le t \le T_b \\ 0, & \text{en otro lado} \end{cases} \tag{6.88}$$

donde $i = 1, 2$. En forma correspondiente, los coeficientes s_{ij} para $i = 1, 2$ y $j = 1, 2$ se definen por medio de

$$\begin{aligned} s_{ij} &= \int_0^{T_b} s_i(t)\phi_j(t)\, dt \\ &= \int_0^{T_b} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_i t) \sqrt{\frac{2}{T_b}} \cos(2\pi f_i t)\, dt \\ &= \begin{cases} \sqrt{E_b}, & i = j \\ 0, & i \ne j \end{cases} \end{aligned} \tag{6.89}$$

De ese modo, a diferencia de la PSK binaria coherente, un sistema FSK binario coherente se caracteriza por tener un espacio de señales que es bidimensional (es decir, $N = 2$) con dos puntos de mensaje (es decir, $M = 2$), como se muestra en la figura 6.25. Los dos puntos de mensaje se definen

$$\mathbf{s}_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix} \tag{6.90}$$

y

$$\mathbf{s}_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix} \tag{6.91}$$

con la distancia euclidiana entre ellos igual a $\sqrt{2E_b}$. La figura 6.25 incluye también un par de insertos, que muestran formas de onda representativas de señales $s_1(t)$ y $s_2(t)$.

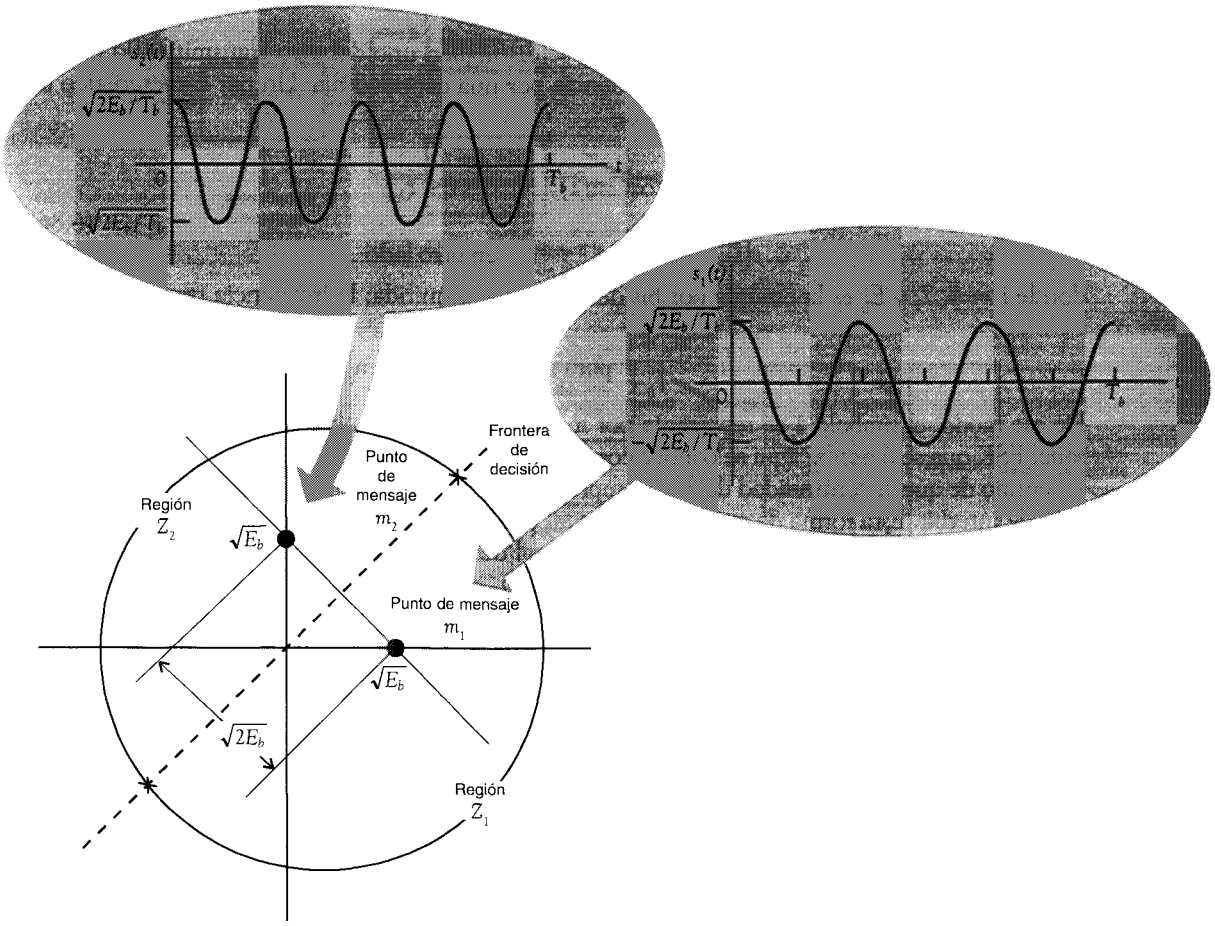


FIGURA 6.25 Diagrama del espacio de señales para sistema FSK binaria. El diagrama incluye también dos insertos que se muestran en formas de onda de ejemplo de dos señales moduladas $s_1(t)$ y $s_2(t)$.

Probabilidad de error de la FSK binaria

El vector de observación **x** tiene dos elementos x_1 y x_2 que se definen, respectivamente, mediante

$$x_1 = \int_0^{T_b} x(t)\phi_1(t)\,dt \tag{6.92}$$

y

$$x_2 = \int_0^{T_b} x(t)\phi_2(t)\,dt \tag{6.93}$$

donde $x(t)$ es la señal recibida, cuya forma depende de cuál fue el símbolo que se transmitió. Debido a que se transmitió el símbolo 1, $x(t)$ es igual a $s_1(t) + w(t)$, donde $w(t)$ es la función de muestreo de un proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$. Si, por otra parte, se transmitió el símbolo 0, $x(t)$ es igual a $s_2(t) + w(t)$.

Ahora bien, aplicando la regla de decisión de la ecuación (5.59), encontramos que el espacio de observación se divide en dos regiones de decisión, marcadas como Z_1 y Z_2 en la figura 6.25. La frontera de decisión, que separa a la región Z_1 de la región Z_2, es el bisector perpendicular de la

línea que une los dos puntos de mensaje. El receptor decide a favor del símbolo 1 si el punto de la señal recibida representado por el vector de observación **x** cae dentro de la región Z_1. Esto ocurre cuando $x_1 > x_2$. Si, en cambio, tenemos que $x_1 < x_2$, el punto de la señal recibida cae dentro de la región Z_2, y el receptor decide a favor del símbolo 0. En la frontera de decisión tenemos $x_1 = x_2$, en cuyo caso el receptor realiza una suposición aleatoria a favor del símbolo 1 o del 0.

Defina una nueva variable aleatoria gaussiana Y cuyo valor de muestreo y es igual a la diferencia entre x_1 y x_2, es decir,

$$y = x_1 - x_2 \tag{6.94}$$

El valor medio de la variable aleatoria Y depende de cuál símbolo binario se transmitió. Debido a que se transmite el símbolo 1, las variables aleatorias gaussianas X_1 y X_2, cuyos valores de muestreo se denotan mediante x_1 y x_2, tienen valores medios iguales a $\sqrt{E_b}$ y cero, respectivamente. De modo correspondiente, la media condicional de la variable aleatoria Y, dado que se transmitió el símbolo 1, es

$$\begin{aligned} E[Y|1] &= E[X_1|1] - E[X_2|1] \\ &= +\sqrt{E_b} \end{aligned} \tag{6.95}$$

Por otro lado, en virtud de que se transmitió el símbolo 0, las variables aleatorias X_1 y X_2 tienen valores medios iguales a cero y $\sqrt{E_b}$, respectivamente. De manera similar, la media condicional de la variable Y, dado que se transmitió el símbolo 0, es

$$\begin{aligned} E[Y|0] &= E[X_1|0] - E[X_2|0] \\ &= -\sqrt{E_b} \end{aligned} \tag{6.96}$$

La varianza de la variable aleatoria Y es independiente de cuál símbolo binario se transmitió. Puesto que las variables aleatorias X_1 y X_2 son independientes estadísticamente, cada una con una varianza igual a $N_0/2$, se concluye que

$$\begin{aligned} \text{var}[Y] &= \text{var}[X_1] + \text{var}[X_2] \\ &= N_0 \end{aligned} \tag{6.97}$$

Supóngase que sabemos que se transmitió el símbolo 0. La función de densidad de probabilidad condicional de la variable aleatoria Y está dada entonces por

$$f_Y(y|0) = \frac{1}{\sqrt{2\pi N_0}} \exp\left[-\frac{(y+\sqrt{E_b})^2}{2N_0}\right] \tag{6.98}$$

Puesto que la condición $x_1 > x_2$, o equivalentemente $y > 0$, corresponde a un receptor que toma una decisión a favor del símbolo 1, deducimos que la probabilidad condicional de error, dado que se transmitió el símbolo 0, es

$$\begin{aligned} p_{10} &= P(y > 0 \,|\, \text{se envió el símbolo 0}) \\ &= \int_0^\infty f_Y(y|0)\, dy \\ &= \frac{1}{\sqrt{2\pi N_0}} \int_0^\infty \exp\left[-\frac{(y+\sqrt{E_b})^2}{2N_0}\right] dy \end{aligned} \tag{6.99}$$

Haciendo

$$\frac{y+\sqrt{E_b}}{\sqrt{2N_0}} = z \tag{6.100}$$

En ese caso, cambiando la variable de integración de y a z, es posible reescribir la ecuación (6.99) del modo siguiente:

$$p_{10} = \frac{1}{\sqrt{\pi}} \int_{\sqrt{E_b/2N_0}}^{\infty} \exp(-z^2)\, dz$$
$$= \frac{1}{2} \operatorname{ferc}\left(\sqrt{\frac{E_b}{2N_0}} \right) \tag{6.101}$$

De modo similar, es posible demostrar que p_{01}, la probabilidad condicional de error debido a que se transmitió el símbolo 1, tiene el mismo valor que en la ecuación (6.101). En consecuencia, promediando p_{10} y p_{01} encontramos que la *probabilidad promedio de error de bits* o, equivalentemente, la *tasa de error de bits para la FSK binaria coherente* (suponiendo símbolos igualmente probables), es

$$P_e = \frac{1}{2} \operatorname{ferc}\left(\sqrt{\frac{E_b}{2N_0}} \right) \tag{6.102}$$

Al comparar las ecuaciones (6.20) y (6.102) vemos que, en un sistema FSK binario coherente, tenemos que duplicar la *relación de la energía de bit a la densidad de ruido*, E_b/N_0, para mantener la misma tasa de error de bits que en un sistema PSK binario coherente. Este resultado concuerda perfectamente con los diagramas del espacio de señales de las figuras 6.3 y 6.25, donde vemos que en un sistema PSK binario la distancia euclidiana entre los dos puntos de mensaje es igual a $2\sqrt{E_b}$, en tanto que en un sistema FSK binario la distancia correspondiente es $\sqrt{2E_b}$. Para una E_b preestablecida, la distancia mínima $d_{\text{mín}}$ en la PSK binaria es, por tanto, $\sqrt{2}$ veces la de la FSK binaria. Recuérdese del capítulo 5 que la probabilidad de error disminuye exponencialmente como $d_{\text{mín}}^2$, de ahí la diferencia entre las fórmulas de las ecuaciones (6.20) y (6.102).

Generación y detección de señales binarias FSK coherentes

Para generar una señal FSK binaria, podemos utilizar el esquema que se presenta en la figura 6.26*a*. La secuencia de datos binarios entrante se aplica primero a un *codificador de nivel activado-desactivado*, a la salida del cual el símbolo 1 se representa mediante una amplitud constante de $\sqrt{E_b}$ volts y el símbolo 0 por medio de cero volts. Utilizando un *inversor* en el canal inferior de la figura 6.26*a*, aseguramos de hecho que cuando tengamos el símbolo 1 a la entrada, se active el oscilador con frecuencia f_1 en el canal superior, en tanto que se desactiva el oscilador con frecuencia f_2 en el canal inferior, con el resultado de que se transmite la frecuencia f_1. Inversamente, cuando tenemos el símbolo 0 a la entrada, se desactiva el oscilador en el canal superior y se activa el del canal inferior, con el resultado de que se transmite la frecuencia f_2. Las dos frecuencias f_1 y f_2 se eligen iguales a diferentes múltiplos enteros de la tasa de bits $1/T_b$, como en la ecuación (6.87).

En el transmisor de la figura 6.26*a*, suponemos que los dos osciladores están sincronizados, de manera que sus salidas satisfacen los requerimientos de las dos funciones básicas ortonormales $\phi_1(t)$ y $\phi_2(t)$, como en la ecuación (6.88). De modo alternativo, es posible utilizar un oscilador de un solo código (controlado por voltaje). En cualquier caso, la frecuencia de la onda modulada se corre con una fase continua, de acuerdo con la onda binaria de entrada.

Para detectar la secuencia binaria original dada la señal ruidosa $x(t)$ que se recibe, podemos utilizar el receptor que se presenta en la figura 6.26*b*, que está compuesto por dos correlatores con una entrada común, los cuales se alimentan con señales de referencia coherentes $\phi_1(t)$ y $\phi_2(t)$ generadas localmente. Las salidas del correlator se sustraen después, una de la otra, y la diferencia que resulta, y, se compara con un umbral de cero volts. Si $y > 0$, el receptor decide a favor de 1. Por otra parte, si $y < 0$, decide a favor de 0. Si y es exactamente cero, el receptor realiza una suposición aleatoria a favor de 1 o 0.

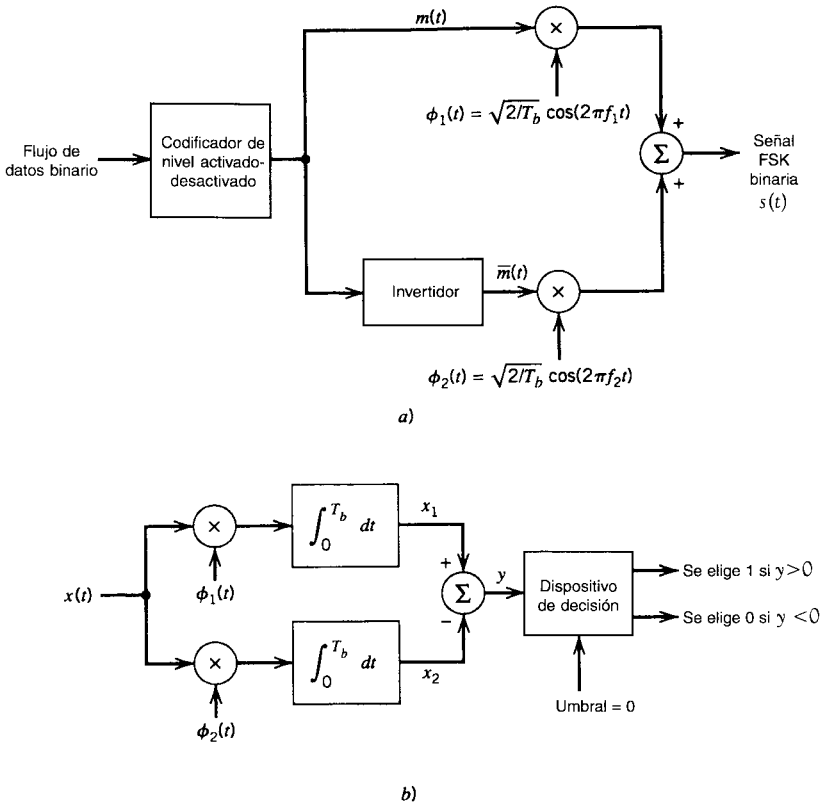


FIGURA 6.26 Diagramas de bloques para *a*) el transmisor FSK binario y *b*) el receptor FSK binario coherente.

Espectros de potencia de señales FSK binarias

Considere el caso de la FSK de Sunde, para el cual las dos frecuencias transmitidas f_1 y f_2 difieren por una cantidad igual a la tasa de bits $1/T_b$, y que su media aritmética es igual a la frecuencia nominal de la portadora f_c; la continuidad de fase se mantiene siempre, incluyendo los tiempos de conmutación de interbit. Podemos expresar esta señal FSK binaria especial de la manera siguiente:

$$s(t) = \sqrt{\frac{2E_b}{T_b}}\cos\left(2\pi f_c t \pm \frac{\pi t}{T_b}\right), \quad 0 \le t \le T_b \tag{6.103}$$

y utilizando una bien conocida identidad trigonométrica, obtener

$$\begin{aligned} s(t) &= \sqrt{\frac{2E_b}{T_b}}\cos\left(\pm\frac{\pi t}{T_b}\right)\cos(2\pi f_c t) - \sqrt{\frac{2E_b}{T_b}}\operatorname{sen}\left(\pm\frac{\pi t}{T_b}\right)\operatorname{sen}(2\pi f_c t) \\ &= \sqrt{\frac{2E_b}{T_b}}\cos\left(\frac{\pi t}{T_b}\right)\cos(2\pi f_c t) \mp \sqrt{\frac{2E_b}{T_b}}\operatorname{sen}\left(\frac{\pi t}{T_b}\right)\operatorname{sen}(2\pi f_c t) \end{aligned} \tag{6.104}$$

En la última línea de la ecuación (6.104), el símbolo más corresponde a transmitir el símbolo 0, y el signo menos indica la transmisión del símbolo 1. Como antes, suponemos que los símbolos 1 y 0 en la onda binaria aleatoria a la entrada del modulador son igualmente probables, y que los símbolos que se transmiten en periodos adyacentes son estadísticamente independientes. En ese caso, con base en

la representación de la ecuación (6.104), podemos establecer las siguientes observaciones relativas a las componentes en fase y en cuadratura de una señal FSK binaria con fase continua:

1. La componente en fase es completamente independiente de la onda binaria de entrada. Es igual a $\sqrt{2E_b/T_b}\cos(\pi t/T_b)$ para todos los valores del tiempo t. La densidad espectral de potencia de esta componente consta, por tanto, de dos funciones delta, ponderadas por el factor $E_b/2T_b$, y que ocurren a $f = \pm 1/2T_b$.
2. La componente en cuadratura se relaciona directamente con la onda binaria de entrada. Durante el intervalo de transmisión $0 \le t \le T_b$, es igual a $-g(t)$ cuando tenemos el símbolo 1, y a $+g(t)$ cuando tenemos el símbolo 0. La función de formación del símbolo $g(t)$ está definida por

$$g(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}}\,\text{sen}\left(\dfrac{\pi t}{T_b}\right), & 0 \le t \le T_b \\ 0, & \text{en otro lado} \end{cases} \tag{6.105}$$

La densidad espectral de energía de esta función de formación del símbolo es igual a

$$\Psi_g(f) = \frac{8E_bT_b\cos^2(\pi T_b f)}{\pi^2(4T_b^2f^2 - 1)^2} \tag{6.106}$$

La densidad espectral de potencia de la componente en cuadratura es igual a $\Psi_g(f)/T_b$. Resulta también claro que las componentes en fase y en cuadratura de la señal FSK binaria son independientes entre sí. Por tanto, la densidad espectral de potencia de la base de la señal FSK de Sunde es igual a la suma de las densidades espectrales de potencia de estas dos componentes, como se señala por medio de

$$S_B(f) = \frac{E_b}{2T_b}\left[\delta\left(f - \frac{1}{2T_b}\right) + \delta\left(f + \frac{1}{2T_b}\right)\right] + \frac{8E_b\cos^2(\pi T_b f)}{\pi^2(4T_b^2f^2 - 1)^2} \tag{6.107}$$

Al sustituir la ecuación (6.107) en la ecuación (6.4), observamos que el espectro de potencia de la señal FSK binaria contiene dos componentes de frecuencia discretas ubicadas en $(f_c + 1/2T_b) = f_1$ y $(f_c - 1/2T_b) = f_2$, con sus potencias promedio agregándose a la mitad de la potencia total de la señal FSK binaria. La presencia de estas dos componentes de frecuencia discretas proporciona un medio para sincronizar el receptor con el transmisor.

Advierta también que la densidad espectral de potencia de banda base de una señal FSK binaria con fase continua cae a la larga como la cuarta potencia inversa de la frecuencia. Esto se establece con facilidad al tomar el límite de la ecuación (6.107) cuando f tiende a infinito. Si, sin embargo, la señal FSK exhibe discontinuidad de fase en los instantes de conmutación de interbit (esto sucede cuando los dos osciladores que aplican las frecuencias f_1 y f_2 operan independientemente uno del otro), la densidad espectral de potencia cae al final como el cuadrado inverso de la frecuencia. Vea el problema 6.23. En consecuencia, una señal FSK con fase continua no produce tanta interferencia fuera de la banda de interés de la señal como una señal FSK con fase discontinua.

En la figura 6.5 hemos graficado los espectros de potencia de banda base de las ecuaciones (6.22) y (6.107). (Para simplificar las cosas, sólo hemos incluido los resultados para la frecuencia positiva.) En ambos casos, $S_B(f)$ se muestra normalizada con respecto a $2E_b$, y la frecuencia se normaliza en relación con la tasa de bits $R_b = 1/T_b$. La diferencia en las tasas de caída de estos espectros puede explicarse con base en la forma del pulso $g(t)$. Tanto más uniforme el pulso, tanto más rápida la caída hacia cero de sus colas espectrales. De este modo, puesto que la FSK binaria (con fase continua) tiene una forma de pulso más uniforme, presenta lóbulos laterales más bajos que la PSK binaria.

Corrimiento mínimo por llaveo

En la detección coherente de una señal FSK binaria, no se explota por completo la información de fase contenida en la señal que se recibe, más que para proporcionar la sincronización entre el receptor y el transmisor. Demostraremos ahora que mediante el uso adecuado de la fase cuando se realiza la detección es posible mejorar considerablemente el desempeño del receptor ante el ruido. Sin embargo, esta mejora se logra a costa de una mayor complejidad del receptor.

Considere una señal de fase continua con corrimiento de frecuencia por llaveo (CPFSK), la cual se define para el intervalo $0 \le t \le T_b$ del modo siguiente:

$$s(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}} \cos[2\pi f_1 t + \theta(0)] & \text{para el símbolo 1} \\ \sqrt{\dfrac{2E_b}{T_b}} \cos[2\pi f_2 t + \theta(0)] & \text{para el símbolo 0} \end{cases} \tag{6.108}$$

donde E_b es la energía por bit de la señal transmitida y T_b es la duración de bit. La fase $\theta(0)$, que denota el valor de la fase en el tiempo $t = 0$, resume la historia pasada del proceso de modulación hasta el tiempo $t = 0$. Las frecuencias f_1 y f_2 se envían en respuesta a los símbolos binarios 1 y 0 que aparecen a la entrada del modulador, respectivamente.

Otra manera útil de representar la señal CPFSK $s(t)$ consiste en expresarla en la forma convencional de una *señal modulada en ángulo* de la manera siguiente:

$$s(t) = \sqrt{\frac{2E_b}{T_b}} \cos[2\pi f_c t + \theta(t)] \tag{6.109}$$

donde $\theta(t)$ es la fase de $s(t)$. Cuando la fase $\theta(t)$ es una función continua del tiempo, encontramos que la propia señal modulada $s(t)$ es también continua todo el tiempo, incluso en los tiempos de conmutación de interbit. La fase $\theta(t)$ de una señal CPFSK aumenta o disminuye linealmente con el tiempo durante cada duración de bit de T_b segundos, de acuerdo con

$$\theta(t) = \theta(0) \pm \frac{\pi h}{T_b} t, \qquad 0 \le t \le T_b \tag{6.110}$$

donde el signo más corresponde a enviar el símbolo 1, y el signo menos al símbolo 0; el parámetro h se va a definir. Mediante la sustitución de la ecuación (6.110) en la (6.109), y la comparación posterior del ángulo de la función con el de la ecuación (6.108), obtenemos el siguiente par de relaciones:

$$f_c + \frac{h}{2T_b} = f_1 \tag{6.111}$$

$$f_c - \frac{h}{2T_b} = f_2 \tag{6.112}$$

Resolviendo las ecuaciones (6.111) y (6.112) para f_c y h, llegamos a

$$f_c = \frac{1}{2}(f_1 + f_2) \tag{6.113}$$

y

$$h = T_b(f_1 - f_2) \tag{6.114}$$

La frecuencia nominal f_c de la portadora es entonces la media aritmética de las frecuencias f_1 y f_2. La diferencia entre las frecuencias f_1 y f_2, normalizada con respecto a la tasa de bit $1/T_b$, define el parámetro adimensional h, el cual se conoce como la *razón de la desviación*.

Enramado de fase

En la ecuación (6.110) encontramos que en el tiempo $t = T_b$,

$$\theta(T_b) - \theta(0) = \begin{cases} \pi h & \text{para el símbolo 1} \\ -\pi h & \text{para el símbolo 0} \end{cases} \tag{6.115}$$

Es decir, enviar el símbolo 1 aumenta la fase de la señal CPFSK $s(t)$ en πh radianes, en tanto que enviar el símbolo 0 la reduce en igual cantidad.

La duración de la fase $\theta(t)$ con el tiempo t sigue una trayectoria que consiste en una secuencia de líneas rectas, cuyas pendientes representan cambios de frecuencia. La figura 6.27 describe todas las trayectorias posibles empezando desde el tiempo $t = 0$. Una gráfica similar que se presenta en la figura 6.27 se conoce como *árbol de fase*. El árbol aclara las transiciones de fase a través de fronteras de intervalo de la secuencia de bits de datos entrante. Además, es evidente, de la figura 6.27, que la fase de una señal CPFSK es un múltiplo impar o par de πh radianes en múltiplos impares o pares de la duración de bits T_b, respectivamente.

El árbol de fase que se describe en la figura 6.27 es una manifestación de la continuidad de fase, la cual es una característica inherente de una señal CPFSK. Para apreciar la noción de la continuidad de fase, vamos a regresar por un momento a la FSK de Sunde, que es un esquema CPFSK como el descrito antes. En ese caso, la razón de desviación h es exactamente la unidad. Por tanto, de acuerdo con la figura 6.27, el cambio de fase sobre un intervalo de bit corresponde a $\pm\pi$ radianes. Sin embargo, un cambio de $+\pi$ radianes es exactamente el mismo que uno de $-\pi$ radianes, módulo 2π. Por tanto, se concluye que en el caso de la FSK de Sunde *no* hay memoria; es decir, conocer cuál cambio particular ocurrió en el intervalo de bit *previo* no ofrece ninguna ayuda en el intervalo de bit *presente*.

En contraste, tenemos una situación por completo diferente cuando a la razón de desviación h se le asigna el valor especial de 1/2. Descubrimos en ese caso que la fase puede encontrar sólo los dos valores $\pm\pi/2$ en múltiplos impares de T_b, y únicamente los dos valores 0 y π en múltiplos pares de T_b, como en la figura 6.28. Esta segunda gráfica recibe el nombre de *enramado de fase*, puesto que un "enramado" es una estructura similar a un árbol con ramas que resurgen. Cada trayectoria de izquierda a derecha a través del enramado de la figura 6.28 corresponde a un entrada de secuencia binaria específica. Por ejemplo, la trayectoria que se presenta remarcada en negro en la figura 6.28 corresponde a la secuencia binaria 1101000 con $\theta(0) = 0$. De aquí en adelante, suponemos que $h = 1/2$.

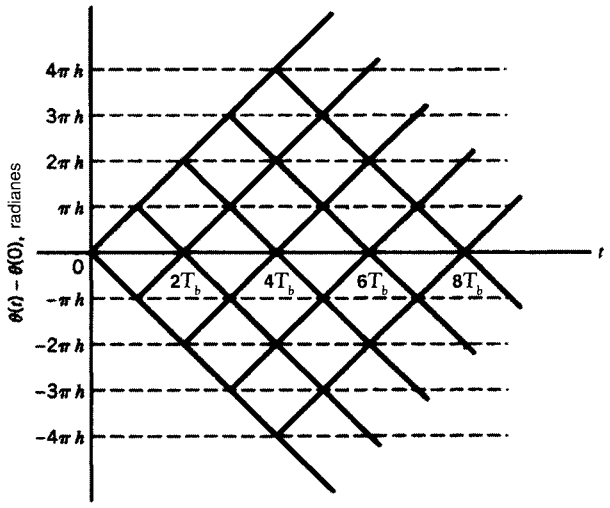


FIGURA 6.27 Árbol de fase.

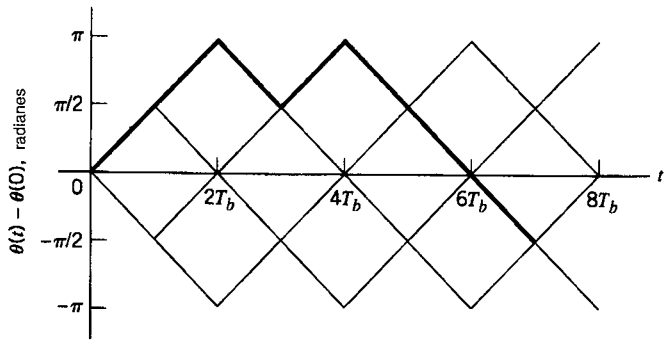


FIGURA 6.28 Enramado de fase; la trayectoria acentuada representa la secuencia 1101000.

Con $h = 1/2$ encontramos, de acuerdo con la ecuación (6.114), que la desviación de frecuencia (es decir, la diferencia entre las dos frecuencias de transmisión de señales f_1 y f_2) es igual a la mitad de la velocidad de bit. Éste es el espaciamiento de frecuencia mínimo que permite que las dos señales FSK que representan 1 y 0, como en la ecuación (6.108), sean coherentemente ortogonales en el sentido de que no interfieren entre sí en el proceso de detección. Por esta causa es que a una señal CPFSK con una razón de desviación de un medio suele denominársele *corrimiento mínimo por llaveo* (MSK).[7]

Diagrama de señales del espacio de la MSK

Mediante el uso de la bien conocida identidad trigonométrica de la ecuación (6.109), es posible expresar la señal CPFSK $s(t)$ en términos de sus componentes en fase y en cuadratura del siguiente modo:

$$s(t) = \sqrt{\frac{2E_b}{T_b}}\cos[\theta(t)]\cos(2\pi f_c t) - \sqrt{\frac{2E_b}{T_b}}\,\text{sen}[\theta(t)]\,\text{sen}(2\pi f_c t) \tag{6.116}$$

Considere primero la componente en fase $\sqrt{2E_b/T_b}\cos[\theta(t)]$. Con la razón de desviación $h = 1/2$, tenemos de la ecuación (6.110) que

$$\theta(t) = \theta(0) \pm \frac{\pi}{2T_b}t, \qquad 0 \le t \le T_b \tag{6.117}$$

donde el signo más corresponde al símbolo 1 y el signo menos al símbolo 0. Un resultado similar se cumple para $\theta(t)$ en el intervalo $-T_b \le t \le 0$, salvo porque el signo algebraico no es necesariamente el mismo en ambos intervalos. Puesto que la fase $\theta(0)$ es 0 o π, dependiendo de la historia pasada del proceso de modulación, encontramos que en el intervalo $-T_b \le t \le T_b$, la polaridad de $\cos[\theta(t)]$ depende exclusivamente de $\theta(0)$, independientemente de la secuencia de unos o ceros transmitida antes o después de $t = 0$. De ese modo, para este intervalo, la componente en fase $s_I(t)$ consistente de un *pulso coseno de medio ciclo* se define de la manera siguiente:

$$\begin{aligned} s_I(t) &= \sqrt{\frac{2E_b}{T_b}}\cos[\theta(t)] \\ &= \sqrt{\frac{2E_b}{T_b}}\cos[\theta(t)]\cos\left(\frac{\pi}{2T_b}t\right) \\ &= \pm\sqrt{\frac{2E_b}{T_b}}\cos\left(\frac{\pi}{2T_b}t\right), \qquad -T_b \le t \le T_b \end{aligned} \tag{6.118}$$

donde el signo más corresponde a $\theta(0) = 0$ y el signo menos a $\theta(0) = \pi$. De manera similar, podemos demostrar que en el intervalo $0 \le t \le 2T_b$ la componente en cuadratura $s_Q(t)$ consiste en un *pulso seno de medio ciclo*, cuya polaridad depende sólo de $\theta(T_b)$, como indica

$$\begin{aligned} s_Q(t) &= \sqrt{\frac{2E_b}{T_b}} \operatorname{sen}[\theta(t)] \\ &= \sqrt{\frac{2E_b}{T_b}} \operatorname{sen}[\theta(T_b)] \operatorname{sen}\left(\frac{\pi}{2T_b}t\right) \\ &= \pm\sqrt{\frac{2E_b}{T_b}} \operatorname{sen}\left(\frac{\pi}{2T_b}t\right), \qquad 0 \le t \le 2T_b \end{aligned} \tag{6.119}$$

donde el signo más corresponde a $\theta(T_b) = \pi/2$ y el signo menos a $\theta(T_b) = -\pi/2$.

De acuerdo con la explicación anterior y en vista de que los estados de fase $\theta(0)$ y $\theta(T_b)$ pueden asumir uno de dos valores posibles, es viable que surja una de cuatro posibilidades, como se describe a continuación:

- La fase $\theta(0) = 0$ y $\theta(T_b) = \pi/2$, correspondiendo a la transmisión del símbolo 1.
- La fase $\theta(0) = \pi$ y $\theta(T_b) = \pi/2$, correspondiendo a la transmisión del símbolo 0.
- La fase $\theta(0) = \pi$ y $\theta(T_b) = -\pi/2$ (o, de manera equivalente, $3\pi/2$ módulo 2π), que corresponde a la transmisión del símbolo 1.
- La fase $\theta(0) = 0$ y $\theta(T_b) = -\pi/2$, que corresponde a la transmisión del símbolo 0.

Esto, a su vez, significa que la propia señal MSK puede tomar cualesquiera de cuatro formas posibles, dependiendo de los valores de $\theta(0)$ y $\theta(T_b)$.

De acuerdo con la expansión de la ecuación (6.116), deducimos que las funciones básicas ortonormales $\phi_1(t)$ y $\phi_2(t)$ para MSK se definen mediante un par de portadoras en cuadratura moduladas senoidalmente:

$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos\left(\frac{\pi}{2T_b}t\right)\cos(2\pi f_c t), \qquad 0 \le t \le T_b \tag{6.120}$$

$$\phi_2(t) = \sqrt{\frac{2}{T_b}} \operatorname{sen}\left(\frac{\pi}{2T_b}t\right)\operatorname{sen}(2\pi f_c t), \qquad 0 \le t \le T_b \tag{6.121}$$

De modo correspondiente, podemos expresar la señal MSK en la forma expandida

$$s(t) = s_1\phi_1(t) + s_2\phi_2(t), \qquad 0 \le t \le T_b \tag{6.122}$$

donde los coeficientes s_1 y s_2 se relacionan con los estados de fase $\theta(0)$ y $\theta(T_b)$, respectivamente. Para evaluar s_1 integramos el producto $s(t)\phi_1(t)$ entre los límites $-T_b$ y T_b, como se indica mediante

$$\begin{aligned} s_1 &= \int_{-T_b}^{T_b} s(t)\phi_1(t)\,dt \\ &= \sqrt{E_b}\cos[\theta(0)], \qquad -T_b \le t \le T_b \end{aligned} \tag{6.123}$$

De modo similar, para evaluar s_2 integramos el producto $s(t)\phi_2(t)$ entre los límites 0 y $2T_b$, indicado como

$$\begin{aligned} s_2 &= \int_0^{2T_b} s(t)\phi_2(t)\,dt \\ &= -\sqrt{E_b}\operatorname{sen}[\theta(T_b)], \qquad 0 \le t \le 2T_b \end{aligned} \tag{6.124}$$

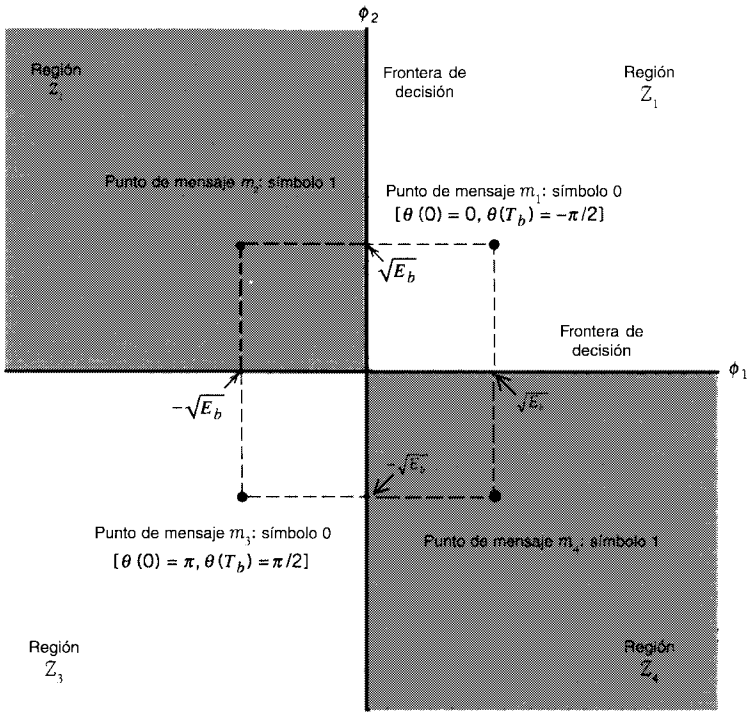


FIGURA 6.29 Diagrama del espacio de señales para el sistema MSK.

Advierta que en las ecuaciones (6.123) y (6.124):

- Ambas integrales se evalúan en un intervalo de tiempo igual al doble de la duración de bit.
- Tanto el límite superior como el inferior de la integración del producto que se emplean para evaluar el coeficiente s_1 se corren en una duración de bit T_b con respecto a aquellos que se utilizan para evaluar el coeficiente s_2.
- El intervalo de tiempo $0 \leq t \leq T_b$, para el cual se definen los estados de fase $\theta(0)$ y $\theta(T_b)$, es común para ambas integrales.

Por tanto, la constelación de señales para una señal MSK es bidimensional (es decir, $N = 2$), con cuatro puntos de mensaje posibles (es decir, $M = 4$), como se ilustra en la figura 6.29. Las coordenadas de los puntos de mensaje son las siguientes en una dirección contraria a la de las manecillas del reloj: $(+\sqrt{E_b}, +\sqrt{E_b}), (-\sqrt{E_b}, +\sqrt{E_b}), (-\sqrt{E_b}, -\sqrt{E_b})$, y $(+\sqrt{E_b}, -\sqrt{E_b})$. Los valores posibles de $\theta(0)$ y $\theta(T_b)$, correspondientes a estos cuatro puntos de mensaje, se incluyen también en la figura 6.29. En consecuencia, el diagrama del espacio de señales de MSK es similar al de la QPSK en que ambos tienen cuatro puntos de mensaje. Sin embargo, difieren de una manera sutil que debe observarse con cuidado: en la QPSK el símbolo transmitido se representa por medio de uno de los cuatro puntos de mensaje, en tanto que en la MSK uno de los dos puntos de mensaje se utiliza para representar el símbolo transmitido en cualquier tiempo, dependiendo del valor de $\theta(0)$.

La tabla 6.5 presenta un resumen de los valores de $\theta(0)$ y $\theta(T_b)$, así como los valores correspondientes de s_1 y s_2 que se calculan para los intervalos de tiempo $-T_b \leq t \leq T_b$ y $0 \leq t \leq 2T_b$, respectivamente. La primera columna en esta tabla indica si el símbolo 1 o el símbolo 0 se enviaron en el intervalo $0 \leq t \leq T_b$. Advierta que las coordenadas de los puntos de mensaje, s_1 y s_2, tienen signos opuestos cuando se envía el símbolo 1 en este intervalo, aunque el signo es igual cuando se envía el símbolo 0. Por tanto, para una secuencia determinada de datos de

TABLA 6.5 *Caracterización del espacio de señales de la MSK*

Símbolo binario transmitido, $0 \le t \le T_b$	*Estados de fase (radianes)*		*Coordenadas de puntos de mensaje*	
	$\theta(0)$	$\theta(T_b)$	s_1	s_2
0	0	$-\pi/2$	$+\sqrt{E_b}$	$+\sqrt{E_b}$
1	π	$-\pi/2$	$-\sqrt{E_b}$	$+\sqrt{E_b}$
0	π	$+\pi/2$	$-\sqrt{E_b}$	$-\sqrt{E_b}$
1	0	$+\pi/2$	$+\sqrt{E_b}$	$-\sqrt{E_b}$

entrada, podemos utilizar las entradas de la tabla 6.5 para obtener, en un esquema de bit por bit, las dos secuencias de coeficientes que se requieren para escalar $\phi_1(t)$ y $\phi_2(t)$ y determinar a partir de ello la señal MSK $s(t)$.

▶ **EJEMPLO 6.5**

La figura 6.30 muestra las secuencias y formas de onda implicadas en la generación de una señal MSK para una secuencia binaria 1101000. La secuencia binaria de entrada se ilustra en la figura 6.30*a*. Las dos frecuencias de modulación son $f_1 = 5/4T_b$ y $f_2 = 3/4T_b$. Suponiendo que en el tiempo $t = 0$ la fase $\theta(0)$ es cero, la secuencia de estados de fase es como se muestra en la figura 6.30, módulo 2π. Las polaridades que se utilizan para escalar las funciones de tiempo $\phi_1(t)$ y $\phi_2(t)$ se muestran en las líneas superiores de las figuras 6.30*b* y 6.30*c*. Advierta que estas dos secuencias están *descompensadas* una con respecto a la otra por un intervalo igual a la duración de bit T_b. Las formas de onda de las dos componentes resultantes de $s(t)$, a saber $s_1\phi_1(t)$ y $s_2\phi_2(t)$, se muestran también en las figuras 6.30*b* y 6.30*c*. Sumando estas dos formas de onda moduladas, obtenemos la señal MSK $s(t)$ que se desea y que se presenta en la figura 6.30*d*. ◀

Probabilidad de error de la MSK

En el caso de un canal AWGN, la señal recibida está dada por

$$x(t) = s(t) + w(t)$$

donde $s(t)$ es la señal MSK transmitida y $w(t)$ es la función de muestreo de un proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$. Para decidir si el símbolo 1 o el símbolo 0 se transmitieron en el intervalo $0 \le t \le T_b$, por ejemplo, tenemos que establecer un procedimiento para el uso de $x(t)$ a fin de detectar los estados de fase $\theta(0)$ y $\theta(T_b)$. Para la detección óptima de $\theta(0)$, determinamos primero la proyección de la señal recibida $x(t)$ sobre la señal de referencia $\phi_1(t)$ en el intervalo $-T_b \le t \le T_b$, obteniendo

$$\begin{aligned} x_1 &= \int_{-T_b}^{T_b} x(t)\phi_1(t)\, dt \\ &= s_1 + w_1, \qquad -T_b \le t \le T_b \end{aligned} \tag{6.125}$$

donde s_1 es como la define la ecuación (6.123) y w_1 es el valor de muestreo de una variable aleatoria gaussiana de media cero y varianza de $N_0/2$. A partir del diagrama del espacio de señales de la figura 6.29, observamos que si $x_1 > 0$, el receptor elige la estimación $\hat{\theta}(0) = 0$. Por otra parte, si $x_1 < 0$, elige la estimación $\hat{\theta}(0) = \pi$.

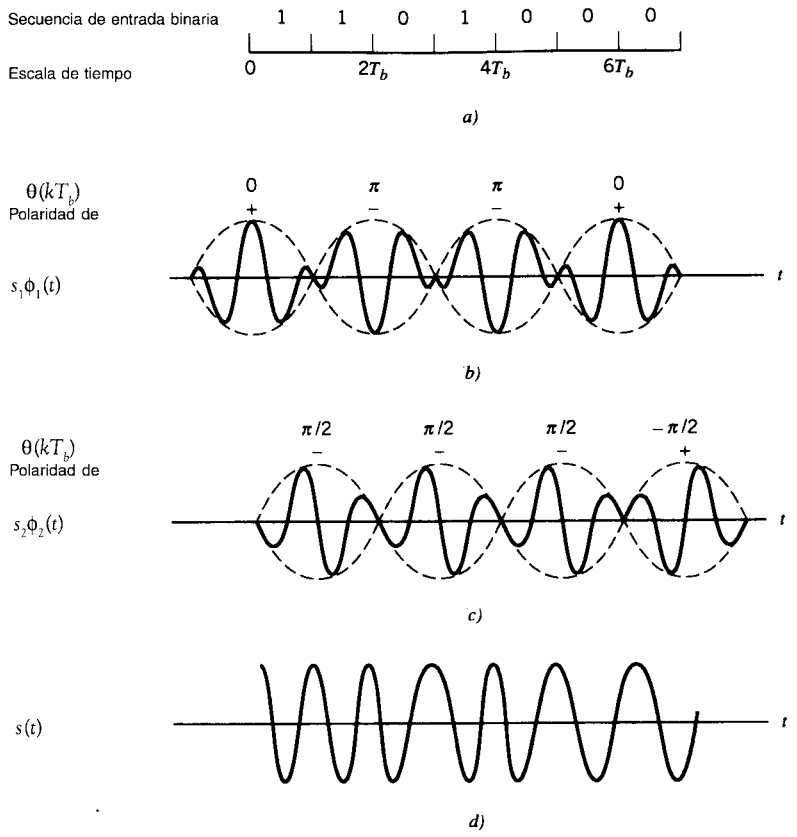


FIGURA 6.30 *a*) Secuencia binaria de entrada. *b*) Forma de onda de la función de tiempo escalada $s_1\phi_1(t)$. *c*) Forma de onda de función de tiempo escalada $s_2\phi_2(t)$. *d*) Forma de onda de la señal MSK $s(t)$ obtenida al sumar $s_1\phi_1(t)$ y $s_2\phi_2(t)$ sobre un esquema de bit por bit.

De manera similar, en la detección óptima de $\theta(T_b)$, determinamos la proyección de la señal recibida $x(t)$ sobre la segunda señal de referencia $\phi_2(t)$ en el intervalo $0 \le t \le 2T_b$, obteniendo

$$\begin{aligned} x_2 &= \int_0^{2T_b} x(t)\phi_2(t)\,dt \\ &= s_2 + w_2, \quad 0 \le t \le 2T_b \end{aligned} \tag{6.126}$$

donde s_2 es como se define en la ecuación (6.124) y w_2 es el valor de muestreo de otra variable aleatoria gaussiana independiente de media cero y varianza $N_0/2$. Refiriéndonos de nuevo al diagrama del espacio de las señales de la figura 6.29, observamos que si $x_2 > 0$, el receptor elige la estimación $\hat{\theta}(T_b) = -\pi/2$. Si, por otro lado $x_2 < 0$, elige la estimación $\hat{\theta}(T_b) = \pi/2$.

Para reconstruir la secuencia binaria original, intercalamos los dos conjuntos anteriores de decisiones de fase, como se describe a continuación (véase la tabla 6.5):

- Si tenemos las estimaciones $\hat{\theta}(0) = 0$ y $\hat{\theta}(T_b) = -\pi/2$, o si tenemos alternativamente las estimaciones $\hat{\theta}(0) = \pi$ y $\hat{\theta}(T_b) = \pi/2$, el receptor toma una decisión a favor del símbolo 0.

- Si tenemos las estimaciones $\hat{\theta}(0) = \pi$ y $\hat{\theta}(T_b) = -\pi/2$, o tenemos de manera alternativa las estimaciones $\hat{\theta}(0) = 0$ y $\hat{\theta}(T_b) = \pi/2$, el receptor toma una decisión a favor del símbolo 1.

De ese modo, al referirnos al diagrama del espacio de señales de la figura 6.29, vemos que la decisión tomada por el receptor se encuentra entre los puntos de mensaje m_1 y m_3 para el símbolo 1, o entre los puntos de mensaje m_2 y m_4 para el símbolo 0. Las decisiones correspondientes a si $\theta(0)$ es 0 o π si $\theta(T_b)$ es $-\pi/2$ o $+\pi/2$ (es decir, las decisiones de bit) se ejecutan *alternativamente* en los canales I y Q del receptor, con cada canal examinando la señal de entrada durante $2T_b$ segundos. La señal de los *demás bits* no interfiere con la decisión del receptor para un bit determinado en cualquier canal. El receptor comete un error cuando el canal I asigna el valor incorrecto a $\theta(0)$ o el canal Q asigna el valor incorrecto a $\theta(T_b)$. Por tanto, utilizando las caracterizaciones estadísticas de las salidas x_1 y x_2 del integrador de producto de estos dos canales, definidas por las ecuaciones (6.125) y (6.126), respectivamente, encontramos de inmediato que la tasa de error de bits para la MSK coherente está dada por

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{6.127}$$

que es exactamente la misma que para la PSK y la QPSK binarias. Sin embargo, es importante advertir que este buen desempeño es el resultado de la detección de la señal MSK que se está realizando en el receptor con base en las observaciones sobre $2T_b$ segundos.

Generación y detección de señales MSK

Considere a continuación la generación y demodulación de la MSK. La figura 6.31*a* muestra el diagrama de bloques de un transmisor MSK característico. La ventaja de este método de generación de señales MSK es que las variaciones en la velocidad de los datos de entrada casi no afectan la coherencia de la señal y la razón de desviación. Las dos ondas senoidales de entrada, una de frecuencia $f_c = n_c/4T_b$ para algún entero fijo n_c, y la otra de frecuencia $1/4T_b$, se aplican primero en el modulador de producto. Esto origina dos ondas senoidales coherentes en fase a frecuencias f_1 y f_2, que se relacionan con la frecuencia de portadora f_c y la tasa de bits $1/T_b$ mediante las ecuaciones (6.111) y (6.112) para $h = 1/2$. Estas dos ondas senoidales se separan una de otra mediante dos filtros de banda angosta, uno centrado en f_1 y el otro en f_2. Las salidas que se producen en los filtros se combinan después linealmente para producir el par de portadoras en cuadratura o funciones básicas ortonormales $\phi_1(t)$ y $\phi_2(t)$. Por último, $\phi_1(t)$ y $\phi_2(t)$ se multiplican con dos ondas binarias $a_1(t)$ y $a_2(t)$, las cuales tienen una velocidad de bits igual a $1/2T_b$. Estas dos ondas binarias se extraen de la secuencia binaria entrante de la manera que se describe en el ejemplo 6.5.

La figura 6.31*b* muestra el diagrama de bloques de un receptor MSK característico. La señal recibida $x(t)$ se correlaciona con réplicas generadas localmente de las señales coherentes de referencia $\phi_1(t)$ y $\phi_2(t)$. Advierta que en ambos casos el intervalo de integración es $2T_b$ segundos, y que la integración en el canal en cuadratura está retardada T_b segundos con respecto a la correspondiente al canal en fase. Las salidas del correlator del canal en fase y en cuadratura x_1 y x_2 que se producen se comparan cada una con un umbral de cero, y se deducen estimaciones de la fase $\theta(0)$ y $\theta(T_b)$ de la manera que se describió antes. Por último, estas decisiones de fase se intercalan de modo que reconstruyan la secuencia binaria original con la probabilidad promedio mínima del error de símbolo en un canal AWGN.

Espectros de potencia de señales MSK

Como con la señal FSK binaria, suponemos que la onda binaria de entrada es aleatoria, con símbolos 1 y 0 igualmente probables, y que los símbolos transmitidos durante diferentes intervalos de tiempo son estadísticamente independientes. En este caso, hacemos las siguientes observaciones:

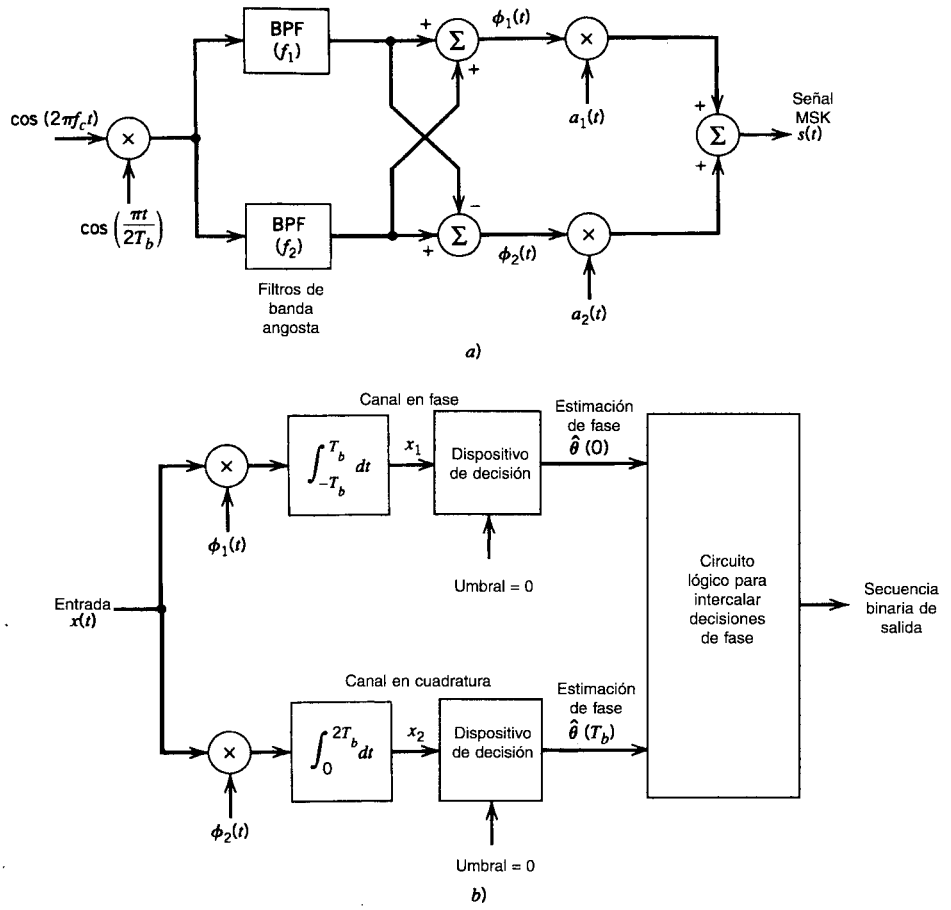


FIGURA 6.31 Diagramas de bloque para *a*) transmisor MSK y *b*) receptor MSK coherente.

1. Dependiendo del valor del estado de fase $\theta(0)$, la componente en fase es igual a $+g(t)$ o $-g(t)$, donde

$$g(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \cos\left(\frac{\pi t}{2T_b}\right), & -T_b \le t \le T_b \\ 0, & \text{en otro caso} \end{cases} \tag{6.128}$$

La densidad espectral de energía de esta función de formación de símbolo es

$$\psi_g(f) = \frac{32E_bT_b}{\pi^2}\left[\frac{\cos(2\pi T_b f)}{16T_b^2 f^2 - 1}\right]^2 \tag{6.129}$$

Por tanto, la densidad espectral de potencia de la componente en fase es igual a $\psi_g(f)/2T_b$.

2. Según el valor del estado de fase $\theta(T_b)$, la componente en cuadratura es igual a $+g(t)$ o $-g(t)$, donde tenemos ahora

$$g(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \operatorname{sen}\left(\frac{\pi t}{2T_b}\right), & 0 \le t \le 2T_b \\ 0, & \text{en otro caso} \end{cases} \tag{6.130}$$

La densidad espectral de energía de esta segunda función de formación del símbolo está dada también por la ecuación (6.129). Por tanto, las componentes en fase y en cuadratura tienen la misma densidad espectral de potencia.

3. Las componentes en fase y en cuadratura de la señal MSK son también estadísticamente independientes. De tal modo, la densidad espectral de potencia de banda base de la señal MSK está determinada por

$$S_B(f) = 2\left[\frac{\psi_g(f)}{2T_b}\right] = \frac{32E_b}{\pi^2}\left[\frac{\cos(2\pi T_b f)}{16T_b^2 f^2 - 1}\right]^2 \tag{6.131}$$

El espectro de potencia de banda base de la ecuación (6.131) se grafica en la figura 6.9, donde el espectro de potencia está normalizado con respecto a $4E_b$ y la frecuencia f se normaliza en relación con la velocidad de bits $1/T_b$. Esta figura incluye también la gráfica correspondiente a la ecuación (6.40) para la señal QPSK, que se consideró antes. Para $f >> 1/T_b$, la densidad espectral de potencia de banda base de la señal MSK disminuye como la cuarta potencia inversa de la frecuencia, en tanto que en el caso de la señal QPSK el descenso corresponde al cuadrado inverso de la frecuencia. Por tanto, la MSK no produce tanta interferencia fuera de la banda de señal de interés como la QPSK. La anterior es una característica deseable de la MSK, sobre todo cuando los sistemas de comunicación digitales operan con una limitación de ancho de banda.

■ MSK CON FILTRADO GAUSSIANO

A partir del estudio detallado del corrimiento mínimo por llaveo (MSK) que acaba de presentarse, podemos resumir las propiedades deseables de la señal MSK del modo siguiente:

- Envolvente constante
- Ancho de banda relativamente angosto
- Desempeño de detección coherente equivalente al de la QPSK

Sin embargo, las características espectrales fuera de banda de las señales MSK, a pesar de ser adecuadas, no satisfacen los severos requisitos de ciertas aplicaciones, como las comunicaciones inalámbricas. Para ilustrar esta limitación, encontramos a partir de la ecuación (6.131) que a $fT_b = 0.5$, la densidad espectral de potencia de banda base de la señal MSK disminuye sólo 10 $\log_{10} 9 = 9.54$ dB por debajo de su valor de banda media. Por tanto, cuando la señal MSK se asigna a un ancho de banda de transmisión de $1/T_b$, la interferencia del canal adyacente del sistema de comunicación inalámbrico utilizando MSK no es lo suficientemente baja para satisfacer los requerimientos prácticos de un ambiente de comunicación multiusuario de este tipo.

Admitiendo que la señal MSK puede generarse mediante la modulación en frecuencia directa de un oscilador controlado por voltaje, es posible superar esta seria limitación de la MSK modificando su espectro de potencia en una *forma compacta*, mientras se mantiene la propiedad de envolvente constante de la señal MSK. Es factible que esta comunicación se logre recurriendo a un filtro pasobajas de premodulación, denominado de aquí en adelante *filtro de formación de pulso* de banda base. Es deseable que este filtro satisfaga las siguientes propiedades:

1. Respuesta en frecuencia con ancho de banda angosto y características de corte definidas.
2. Respuesta al impulso con sobrepaso relativamente bajo.

3. Evolución de un enramado de fase donde la fase portadora de la señal modulada asume los dos valores $\pm\pi/2$ en múltiplos impares de T_b y los dos valores de 0 y π en los múltiplos pares de T_b, como en la MSK.

La condición 1 es necesaria para suprimir las componentes de alta frecuencia de la señal transmitida. La condición 2 evita las desviaciones excesivas en la frecuencia instantánea de la señal de FM. Por último, la condición 3 asegura que la señal de FM modificada puede detectarse coherentemente de la misma manera que la señal MSK, o que puede detectarse no coherentemente como una simple señal FKS binaria.

Es posible conseguir estas propiedades deseables al hacer pasar un flujo de datos binarios sin retorno a cero (NRZ) a través de un filtro de formación de pulso de banda base cuya respuesta al impulso (y de igual manera su respuesta en frecuencia) está definida por una función *gaussiana.* El método resultante de la modulación de frecuencia binaria se denomina de manera natural *MSK de filtrado gaussiano* o sólo GMSK.[8]

Sea W el *ancho de banda de banda base de 3 dB* del filtro de formación de pulso. Podemos definir entonces la función de transferencia $H(f)$ y la respuesta al impulso $h(t)$ del filtro de formación de pulso, respectivamente, de la manera siguiente:

$$H(f) = \exp\left(-\frac{\log 2}{2}\left(\frac{f}{W}\right)^2\right) \tag{6.132}$$

y

$$h(t) = \sqrt{\frac{2\pi}{\log 2}} W \exp\left(-\frac{2\pi^2}{\log 2} W^2 t^2\right) \tag{6.133}$$

La respuesta de este filtro gaussiano a un pulso rectangular de amplitud unitaria y duración T_b (centrada en el origen) está dada por (vea el problema 6.28)

$$\begin{aligned} g(t) &= \int_{-T_b/2}^{T_b/2} h(t-\tau)\, d\tau \\ &= \sqrt{\frac{2\pi}{\log 2}} W \int_{-T_b/2}^{T_b/2} \exp\left(-\frac{2\pi^2}{\log 2} W^2 (t-\tau)^2\right) d\tau \end{aligned} \tag{6.134}$$

que puede expresarse como la diferencia de dos funciones de error complementarias, como se indica mediante

$$g(t) = \frac{1}{2}\left[\operatorname{ferc}\left(\pi\sqrt{\frac{2}{\log 2}} W T_b \left(\frac{t}{T_b} - \frac{1}{2}\right)\right) - \operatorname{ferc}\left(\pi\sqrt{\frac{2}{\log 2}} W T_b \left(\frac{t}{T_b} + \frac{1}{2}\right)\right)\right] \tag{6.135}$$

La respuesta al pulso $g(t)$ constituye el *pulso de formación de frecuencia* del modulador GMSK, con el *producto tiempo-ancho de banda* adimensional WT_b desempeñando la función de un parámetro de diseño.

El pulso de formación de frecuencia $g(t)$, de acuerdo a como se define en la ecuación (6.135), es no casual en el sentido de que es distinto de cero para $t < -T_b/2$, donde $t = -T_b/2$ es el tiempo al cual se aplica el pulso rectangular de entrada (ubicado simétricamente en torno al origen) en el filtro gaussiano. Para una respuesta causal, $g(t)$ debe truncarse y correrse en el tiempo. La figura 6.32 presenta gráficas de $g(t)$, que se han truncado en $t = \pm 2.5T_b$ y se han corrido después en el tiempo en $2.5T_b$. Las gráficas que se muestran aquí corresponden a $WT_b = 0.2$, 0.25 y 0.3. Advierta que cuando se reduce WT_b, la dispersión del tiempo del pulso de formación de frecuencia aumenta de modo correspondiente.

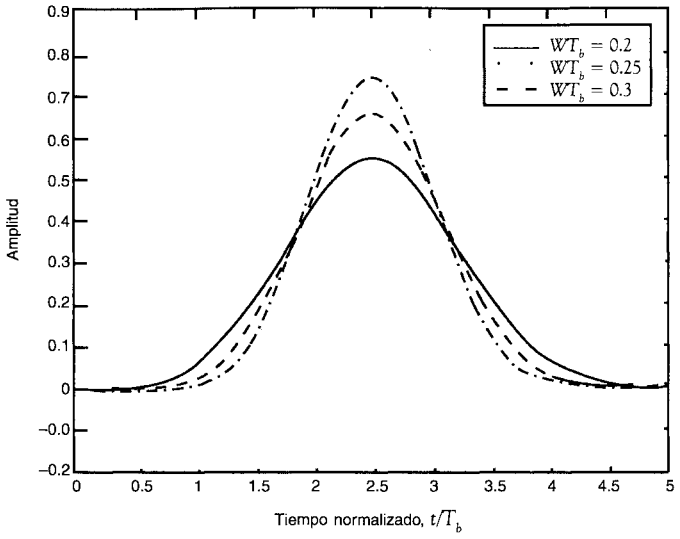


FIGURA 6.32 Pulso de formación de frecuencia $g(t)$ de la ecuación (6.135) corrida en el tiempo en $2.5T_b$ y truncada en $\pm 2.5T_b$ para un producto tiempo-ancho de banda WT_b variable.

La figura 6.33 muestra los espectros de potencia calculados por computadora de señales MSK (expresados en decibeles) en función de la diferencia de frecuencia normalizada $(f - f_c)T_b$, donde f_c es la frecuencia de banda media y T_b es la duración de bit.[9] Los resultados graficados en la figura 6.33 corresponden a valores variables del producto tiempo-ancho de banda WT_b. Con base en esta figura es posible señalar las siguientes observaciones:

- La curva para la condición límite $WT_b = \infty$ corresponde al caso de la MSK ordinaria.
- Cuando WT_b es menor que la unidad, en forma creciente la potencia transmitida se concentra dentro de la banda de paso de la señal GMSK.

Una característica indeseable de la GMSK es que el procesamiento de datos binarios sin retorno a cero por medio de un filtro gaussiano origina una señal moduladora que ya no está confinada a un solo intervalo de bit, como en el caso de la MSK ordinaria, lo cual es claro de inmediato en la figura 6.32. Enunciado de otra manera, las colas de la respuesta al impulso gaussiano del filtro de formación de pulso provocan que la señal moduladora se disperse hacia intervalos de símbolo adyacentes. El resultado neto es la generación de la *interferencia de intersímbolo*, el grado de la cual aumenta al disminuir WT_b. A la luz de esta observación y de la que se hizo en la figura 6.33 sobre los espectros de potencia de las señales GMSK, podemos afirmar que la elección del producto tiempo-ancho de banda WT_b ofrece un intercambio entre la compactación espectral y la pérdida de desempeño.

Para investigar el aspecto de la degradación del desempeño, considere la probabilidad de error P_e de la GMSK utilizando detección coherente en la presencia de ruido blanco gaussiano aditivo. Admitiendo que la GMSK es un tipo especial de modulación en frecuencia binaria, es posible expresar P_e mediante la fórmula empírica

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{\alpha E_b}{2N_0}}\right) \tag{6.136}$$

donde, como antes, E_b es la energía de la señal por bit y $N_0/2$ es la densidad espectral del ruido. El factor α es una constante cuyo valor depende del producto tiempo-ancho de banda WT_b. Comparando la fórmula de la ecuación (6.136) para la GMSK, con la de la ecuación (6.127) para la MSK

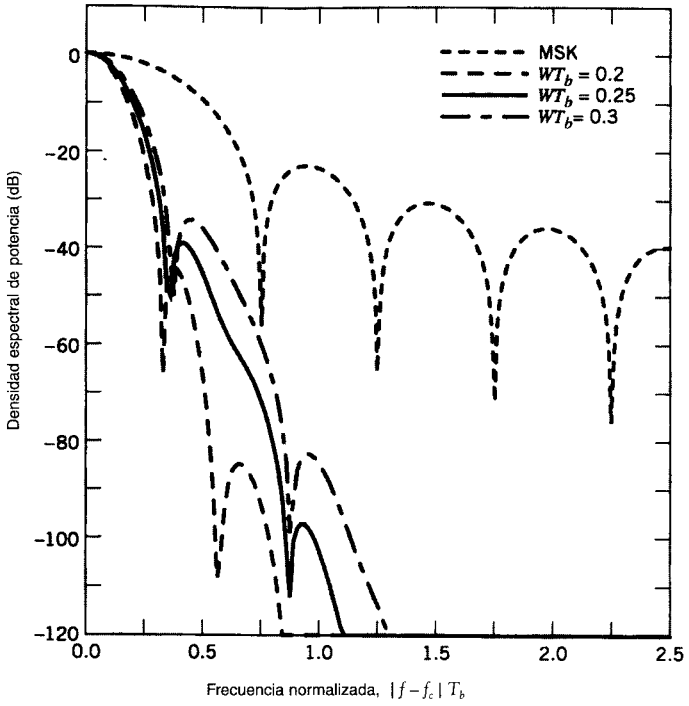


FIGURA 6.33 Espectro de potencia de las señales MSK y GMSK para producto variable tiempo-ancho de banda. (Reproducido con autorización del doctor Gordon Stüber, Georgia Tech.)

ordinaria, podemos considerar a $10\log_{10}(\alpha/2)$, expresado en decibeles, como una medida de la degradación del desempeño de la GMSK (con WT_b preestablecido), en comparación con la MSK ordinaria. La figura 6.34 representa el valor calculado en computadora de $10\log_{10}(\alpha/2)$ en función de WT_b. Para la MSK ordinaria tenemos $WT_b = \infty$, en cuyo caso la ecuación (6.136) con $\alpha = 2$ asume exactamente la misma forma que la ecuación (6.127) y no hay degradación en el desempeño, lo cual confirma la figura 6.34. Para una GMSK con $WT_b = 0.3$ encontramos, de acuerdo con

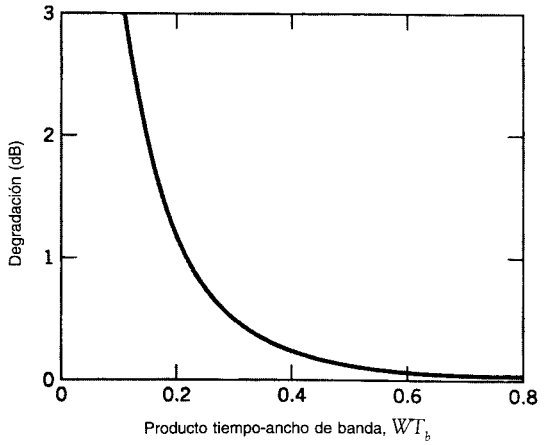


FIGURA 6.34 Degradación teórica E_b/N_0 de la GMSK para el producto tiempo-ancho de banda variable. (Tomada de Murata e Hirade, con permiso del IEEE.)

la figura 6.34, que existe una degradación en el desempeño de aproximadamente 0.46 dB, lo cual corresponde a $(\alpha/2) = 0.9$. Esta degradación en el desempeño es un pequeño precio que debe pagarse para obtener la compactación espectral altamente deseable de la señal GMSK.

EJEMPLO 6.6 GMSK para comunicaciones inalámbricas GSM

Una aplicación importante de la GMSK se encuentra en un sistema de comunicación inalámbrico estandarizado ampliamente conocido como GSM, que es un sistema de acceso múltiple por división de tiempo que se explica en el capítulo 8. En esta aplicación, el producto tiempo-ancho de banda WT_b de la GMSK se estandariza en 0.3, lo cual proporciona el mejor compromiso entre la ocupación del ancho de banda incrementado y la resistencia a la interferencia de cocanal. Noventa y nueve por ciento de la potencia de radio frecuencia (RF) de señales GMSK especificadas de esa manera se encuentran confinadas en un ancho de banda de 250 kHz, lo cual significa que, para todos los fines prácticos, los lóbulos laterales son virtualmente cero fuera de esta banda de frecuencia.

El espectro disponible se divide en subcanales de 200 kHz de ancho y cada uno se asigna a un sistema GSM que transmite datos a 271 kb/s. La figura 6.35 describe el espectro de potencia de un subcanal en relación con sus dos subcanales adyacentes. Esta gráfica es la versión pasobanda del espectro de potencia de banda base de la figura 6.33 que corresponde a $WT_b = 0.3$. Con base en la figura 6.35 podemos efectuar la siguiente observación importante: el espectro de potencia RF del subcanal que se presenta sombreado se reduce en una cantidad mayor que 40 dB a las frecuencias de portadora de ambos subcanales adyacentes, lo que quiere decir que el efecto de la interferencia de cocanal es prácticamente ignorable. ◀

FSK M-ARIA

Considere a continuación la versión M-aria de la FSK, para la cual las señales transmitidas se definen por medio de

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[\frac{\pi}{T}(n_c + i)t\right], \qquad 0 \le t \le T \tag{6.137}$$

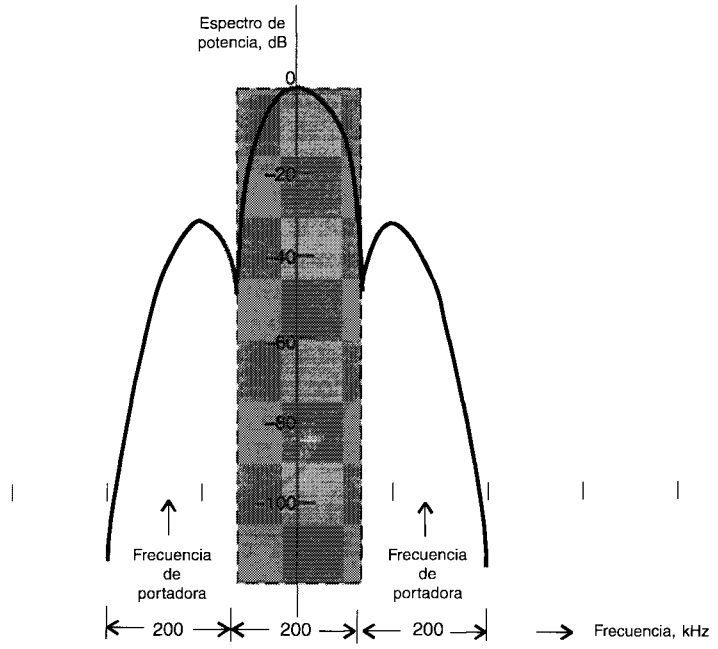


FIGURA 6.35 Espectro de potencia de la señal GMSK para las comunicaciones inalámbricas.

donde $i = 1, 2, \ldots, M$, y la frecuencia de portadora $f_c = n_c/2T$ para algún entero fijo n_c. Los símbolos transmitidos son de igual duración T y tienen igual energía E. Puesto que las frecuencias de las señales individuales están separadas por $1/2T$ Hz, las señales en la ecuación (6.137) son ortogonales; es decir,

$$\int_0^T s_i(t)s_j(t)\,dt = 0, \qquad i \neq j \tag{6.138}$$

Esta propiedad de la FSK M-aria sugiere que es posible utilizar las propias señales transmitidas $s_i(t)$, salvo por la normalización de energía, como un conjunto ortogonal completo de funciones básicas, como se muestra por medio de

$$\phi_i(t) = \frac{1}{\sqrt{E}} s_i(t), \qquad \begin{matrix} 0 \le t \le T \\ i = 1, 2, \ldots, M \end{matrix} \tag{6.139}$$

En consecuencia, la FSK M-aria se describe por medio de un diagrama del espacio de señales M-dimensional.

Para la FSK M-aria coherente, el receptor óptimo consiste en un banco de M correlatores o filtros de acoplamiento, con la $\phi_i(t)$ de la ecuación (6.139) proporcionando las señales de referencia pertinentes. En los tiempos de muestreo $t = kT$, el receptor toma decisiones con base en la salida más grande de filtro de acoplamiento en concordancia con la regla de decodificación de máxima verosimilitud. Sin embargo, una fórmula exacta para la probabilidad del error de símbolo resulta difícil de obtener para un sistema FSK M-ario coherente. A pesar de eso, podemos utilizar la cota de unión de la ecuación (5.96) del capítulo 5 para imponer una cota superior sobre la probabilidad promedio del error de símbolo para la FSK M-aria. De manera específica, observando que la distancia mínima $d_{\text{mín}}$ en la FSK M-aria es igual a $\sqrt{2E}$, el uso de la ecuación (5.96) produce (suponiendo símbolos equiprobables)

$$P_e \le \frac{1}{2}(M-1)\ \text{ferc}\left(\sqrt{\frac{E}{2N_0}}\right) \tag{6.140}$$

Para M fija, esta cota se vuelve cada vez más ajustada cuando aumenta E/N_0. En realidad se vuelve una buena aproximación a P_e para valores de $P_e \le 10^{-3}$. Asimismo, para $M = 2$ (es decir, FSK binaria), la cota de la ecuación (6.140) se convierte en una igualdad.

Espectro de potencias de señales FSK M-arias

El análisis espectral de señales FSK M-arias[10] es mucho más complicado que el correspondiente a las señales PSK M-arias. Un caso de interés particular se presenta cuando las frecuencias asignadas a los multiniveles uniforman el espaciamiento de frecuencia y hacen que la desviación de frecuencia corresponda a $k = 0.5$. Es decir, las M frecuencias de señal están separadas por $1/2T$, donde T es la duración del símbolo. Para $k = 0.5$, la densidad espectral de potencia de banda base de las señales FKS M-arias se grafica en la figura 6.36 para $M = 2, 4, 8$.

Eficiencia del ancho de banda de señales FSK M-arias

Cuando se detectan coherentemente las señales ortogonales de una señal FSK M-aria, las señales adyacentes sólo necesitan estar separadas una de otra por una diferencia de frecuencia $1/2T$ para que así se mantenga la ortogonalidad. Por tanto, es posible definir el ancho de canal requerido para transmitir las señales FSK M-arias como

$$B = \frac{M}{2T} \tag{6.141}$$

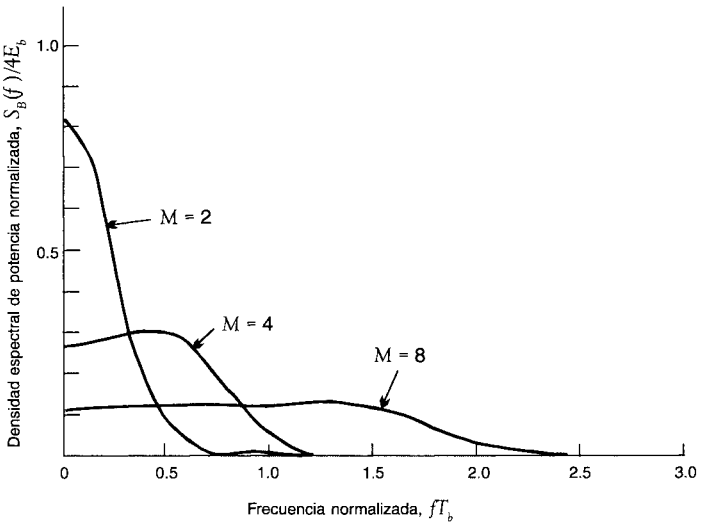


FIGURA 6.36 Espectros de potencia de señales PSK M-arias para M = 2, 4, 8.

Para multiniveles con asignamientos de frecuencias que hacen que el espaciamiento de éstas sea uniforme e igual a $1/2T$, el ancho de banda B de la ecuación (6.141) contiene una fracción considerable de la potencia de la señal. Esto se confirma de inmediato al observar las gráficas espectrales de la potencia de banda base que se ilustran en la figura 6.36. A partir de la ecuación (6.48) recordamos que el periodo del símbolo T es igual a $T_b \log_2 M$. Por tanto, utilizando $R_b = 1/T_b$ podemos redefinir el ancho de banda del canal B para señales FSK M-arias como

$$B = \frac{R_b M}{2 \log_2 M} \tag{6.142}$$

La eficiencia del ancho de banda de las señales M-arias es entonces

$$\rho = \frac{R_b}{B} = \frac{2 \log_2 M}{M} \tag{6.143}$$

La tabla 6.6 proporciona los valores de ρ calculados con base en la ecuación (6.143) para M variable.

Al comparar las tablas 6.4 y 6.6 observamos que al aumentar el número de niveles M se tiende a incrementar la eficiencia del ancho de banda de las señales PSK M-arias, pero que también se tiende a reducir la eficiencia del ancho de banda de las señales FSK M-arias. En otras palabras, las señales PSK M-arias son eficientes espectralmente, en tanto que las FSK M-arias son espectralmente ineficientes.

TABLA 6.6 *Eficiencia del ancho de banda de señales FSK M-arias*

M	2	4	8	16	32	64
ρ (bits/s/Hz)	1	1	0.75	0.5	0.3125	0.1875

6.6 *Detección de señales con fase desconocida*

Hasta este punto de nuestra exposición, hemos supuesto que el receptor está perfectamente sincronizado con el transmisor y que el único deterioro del canal es el ruido. Sin embargo, en la práctica muchas veces se encuentra que aparte de la incertidumbre que resulta del ruido del canal, existe también la que surge de la aleatoriedad de ciertos parámetros de la señal. La causa usual de la incertidumbre es la distorsión en el medio de transmisión. Quizás el parámetro de señal aleatorio más común es la fase portadora, lo cual es cierto en especial en el caso de señales de banda angosta. Por ejemplo, la transmisión por una multiplicidad de trayectorias de longitudes diferentes y variables, o los retardos que varían rápidamente en el medio de propagación desde el receptor hasta el transmisor, quizás ocasionen el cambio de fase de la señal recibida de una manera que no puede seguir el receptor. La sincronización con la fase portadora transmitida quizá sea demasiado costosa en esas condiciones, y el diseñador tal vez elija simplemente descartar la información de la fase en la señal recibida a costa de cierta degradación en el desempeño con respecto al ruido. Un receptor de comunicación digital en el que no se provee la recuperación de fase portadora se dice que será *no coherente*.

■ Receptor cuadrático óptimo

Considere un sistema de comunicación digital binario en el cual la señal transmitida es

$$s_i(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_i t), \qquad \begin{matrix} 0 \le t \le T \\ i = 1, 2 \end{matrix} \tag{6.144}$$

donde E es la energía de la señal, T es la duración del intervalo de duración de señales y la frecuencia de la portadora f_i para el símbolo i es un múltiplo entero de $1/2T$. Se supone que el sistema será no coherente, en cuyo caso la señal recibida para un canal AWGN puede escribirse en la forma

$$x(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_i t + \theta) + w(t), \qquad \begin{matrix} 0 \le t \le T \\ i = 1, 2 \end{matrix} \tag{6.145}$$

donde θ es la fase desconocida de la portadora y $w(t)$ es la función de muestreo del proceso de ruido blanco gaussiano de media cero y densidad espectral de potencia $N_0/2$. En una situación práctica, es realista suponer la falta completa de información previa acerca de θ y tratarla como una variable aleatoria con *distribución uniforme:*

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & -\pi < \theta \le \pi \\ 0, & \text{en otro caso} \end{cases} \tag{6.146}$$

El problema de la detección binaria por resolverse puede establecerse ahora del siguiente modo:

> Dada la señal recibida $x(t)$ y confrontada con la fase desconocida θ de la portadora, diséñese un receptor óptimo para detectar el símbolo s_i representado por la componente $\sqrt{E/2T}\cos(2\pi f_i t + \theta)$ que está contenida en $x(t)$.

Procediendo de una manera similar a la que se describe en las secciones 5.3-5.6, es posible formular la *función de verosimilitud condicional* del símbolo s_i, dada la fase θ de la portadora, como

$$L(s_i(\theta)) = \exp\left(\sqrt{\frac{E}{N_0 T}}\int_0^T x(t)\cos(2\pi f_i t + \theta)\, dt\right) \tag{6.147}$$

Para continuar, tenemos que eliminar la dependencia de $L(s_i(\theta))$ de la fase θ, lo cual se logra integrándola sobre todos los valores posibles de θ. Podemos escribir entonces

$$\begin{aligned} L(s_i) &= \int_{-\pi}^{\pi} L(s_i(\theta)) f_\Theta(\theta)\, d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\sqrt{\frac{E}{N_0 T}}\int_0^T x(t)\cos(2\pi f_i t + \theta)\, dt\right) d\theta \end{aligned} \tag{6.148}$$

Advierta que la dependencia de θ se eliminó al integrar la función de verosimilitud y *no* la función de verosimilitud logarítmica.

Utilizando una bien conocida fórmula trigonométrica, podemos expandir $\cos(2\pi f_i t + \theta)$ como

$$\cos(2\pi f_i t + \theta) = \cos(2\pi f_i t)\cos\theta - \operatorname{sen}(2\pi f_i t)\operatorname{sen}\theta$$

De modo correspondiente, se describe la integral en el exponente de la ecuación (6.148) como

$$\int_0^T x(t)\cos(2\pi f_i t + \theta)\, dt = \cos\theta\int_0^T x(t)\cos(2\pi f_i t)\, dt - \operatorname{sen}\theta\int_0^T x(t)\operatorname{sen}(2\pi f_i t)\, dt \tag{6.149}$$

Defina

$$l_i = \left[\left(-\int_0^T x(t)\cos(2\pi f_i t)\, dt\right)^2 + \left(\int_0^T x(t)\operatorname{sen}(2\pi f_i t)\, dt\right)^2\right]^{1/2} \tag{6.150}$$

$$\beta_i = \tan^{-1}\left(\frac{\int_0^T x(t)\operatorname{sen}(2\pi f_i t)\, dt}{\int_0^T x(t)\cos(2\pi f_i t)\, dt}\right) \tag{6.151}$$

Por tanto, podemos ir un paso adelante y simplificar la ecuación (6.149) a:

$$\begin{aligned} \int_0^T x(t)\cos(2\pi f_i t + \theta)\, dt &= l_i(\cos\theta\cos\beta_i - \operatorname{sen}\theta\operatorname{sen}\beta_i) \\ &= l_i\cos(\theta + \beta_i) \end{aligned} \tag{6.152}$$

Por tanto, al utilizar la ecuación (6.152) en la (6.148), obtenemos

$$\begin{aligned} L(s_i) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\sqrt{\frac{E}{N_0 T}} l_i \cos(\theta + \beta_i)\right) d\theta \\ &= \frac{1}{2\pi}\int_{-\pi+\beta_i}^{\pi+\beta_i} \exp\left(\sqrt{\frac{E}{N_0 T}} l_i \cos\theta\right) d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\sqrt{\frac{E}{N_0 T}} l_i \cos\theta\right) d\theta \end{aligned} \tag{6.153}$$

A partir del apéndice 3 acerca de las funciones de Bessel, reconocemos de inmediato la integral de la ecuación (6.153) como la *función de Bessel modificada de orden cero:*

$$I_0\left(\sqrt{\frac{E}{N_0 T}} l_i\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(\sqrt{\frac{E}{N_0 T}} l_i \cos\theta\right) d\theta \tag{6.154}$$

De ese modo, podemos expresar la función de verosimilitud para el problema de detección de señales descrito aquí en la forma compacta

$$L(s_i) = I_0\left(\sqrt{\frac{E}{N_0 T}}\, l_i\right) \tag{6.155}$$

La *prueba de hipótesis binaria* (es decir, la hipótesis de que la señal $s_1(t)$ o la señal $s_2(t)$ se transmitieron) puede escribirse ahora como

$$I_0\left(\sqrt{\frac{E}{N_0 T}}\, l_1\right) \underset{H_2}{\overset{H_1}{\gtrless}} I_0\left(\sqrt{\frac{E}{N_0 T}}\, l_2\right) \tag{6.156}$$

donde las hipótesis H_1 y H_2 corresponden a las señales $s_1(t)$ y $s_2(t)$, respectivamente. Según el apéndice 3 observamos que la función de Bessel modificada $I(\cdot)$ es monotónicamente creciente. Por tanto, la prueba de hipótesis puede realizarse en términos ya sea de $I_0(\sqrt{E/N_0 T}\, l_i)$ o simplemente l_i. Por conveniencia en la puesta en práctica, la prueba de hipótesis se lleva a cabo en términos de l_i^2 en lugar de l_i, como se indica mediante

$$l_1^2 \underset{H_2}{\overset{H_1}{\gtrless}} l_2^2 \tag{6.157}$$

Un receptor que se basa en la ecuación (6.157) recibe el nombre de *receptor cuadrático*. A la luz de la definición de l_i dada en la ecuación (6.150), la estructura del receptor para el cálculo de l_i es como se indica en la figura 6.37*a*. Observe que la prueba descrita en la ecuación (6.157) resulta independiente de la energía del símbolo E. Por esta razón es que la prueba de hipótesis se dice que será *uniformemente más poderosa* con respecto a la energía del símbolo E.

Dos formas equivalentes del receptor cuadrático

A continuación obtenemos dos formas equivalentes del receptor en cuadratura que se muestra en la figura 6.37*a*. La primera forma se obtiene con facilidad al sustituir cada correlator en la figura 6.37*a* con un correspondiente filtro de acoplamiento equivalente. De ese modo obtenemos la forma alternativa del receptor en cuadratura que se muestra en la figura 6.37*b*. En una rama de este receptor tenemos un filtro de acoplamiento a la señal $\cos(2\pi f_i t)$, y en la otra, un filtro de acoplamiento a $\operatorname{sen}(2\pi f_i t)$, los cuales se definen para el intervalo de tiempo $0 \le t \le T$. Las salidas del filtro se muestrean en el tiempo $t = T$, se elevan al cuadrado y se suman.

Para obtener la segunda forma equivalente de receptor en cuadratura, suponga que tenemos un filtro que está acoplado a $s(t) = \cos(2\pi f_i t + \theta)$ para $0 \le t \le T$. La envolvente de la salida del filtro de acoplamiento evidentemente no se ve afectada por el valor de la fase θ. De ese modo, por conveniencia, podemos simplemente elegir un filtro de acoplamiento con respuesta al impulso $\cos[2\pi f_i(T - t)]$, correspondiente a $\theta = 0$. La salida de un filtro de este tipo en respuesta a una señal recibida $x(t)$ está dada por

$$\begin{aligned} y(t) &= \int_0^T x(\tau)\, \cos\left[2\pi f_i (T - t + \tau)\right] d\tau \\ &= \cos\left[2\pi f_i (T - t)\right] \int_0^T x(\tau)\, \cos(2\pi f_i \tau)\, d\tau - \operatorname{sen}\left[2\pi f_i (T - t)\right] \int_0^T x(\tau)\, \operatorname{sen}(2\pi f_i \tau)\, d\tau \end{aligned} \tag{6.158}$$

La envolvente de la salida del filtro de acoplamiento es proporcional a la raíz cuadrada de la suma de los cuadrados de las integrales en la ecuación (6.158). La envolvente, evaluada en el tiempo $t = T$, es entonces

$$l_i = \left\{\left[\int_0^T x(\tau)\, \cos(2\pi f_i \tau)\, d\tau\right]^2 + \left[\int_0^T x(\tau)\, \operatorname{sen}(2\pi f_i \tau)\, d\tau\right]^2\right\}^{1/2} \tag{6.159}$$

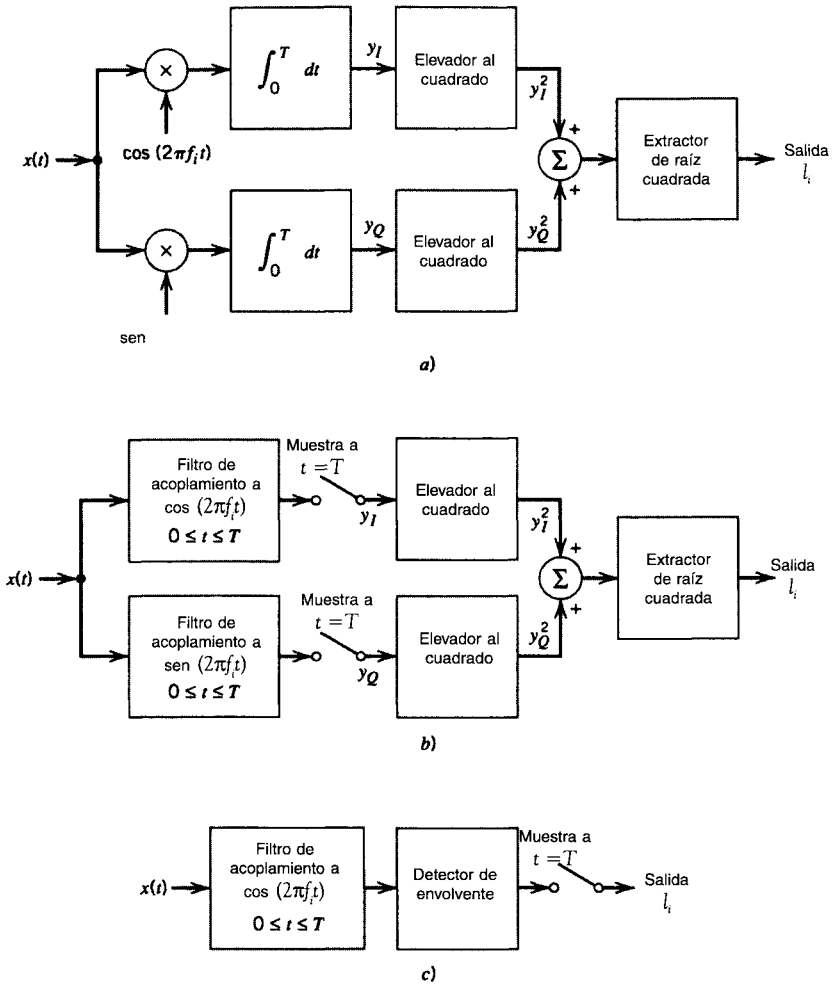


FIGURA 6.37 Receptores no coherentes. *a*) Receptor en cuadratura utilizando correlatores. *b*) Receptor en cuadratura utilizando filtros de acoplamiento. *c*) Filtro de acoplamiento no coherente.

Sin embargo lo anterior es sólo la salida del receptor en cuadratura. Por tanto, la salida (en el tiempo T) de un filtro de acoplamiento a la señal $\cos(2\pi f_i t + \theta)$ de fase arbitraria θ seguida por un detector de envolvente es la misma que la salida correspondiente del receptor en cuadratura de la figura 6.37*a*. Esta forma de receptor se presenta en la figura 6.37*c*. La combinación del filtro de acoplamiento y del detector de envolvente que se muestra en la figura 6.37*c* recibe el nombre de *filtro de acoplamiento no coherente*.

La necesidad de un detector de envolvente que siga al filtro de acoplamiento en la figura 6.37*c* puede justificarse también de manera intuitiva del siguiente modo. La salida de un filtro de acoplamiento a una onda RF rectangular alcanza un pico positivo en el instante de muestreo $t = T$. Si, a pesar de eso, la fase del filtro no está acoplada a la de la señal, es posible que el pico ocurra a un tiempo diferente del instante de muestreo. En realidad, si la fase difiere en 180 grados obtenemos un pico negativo en el instante de muestreo. La figura 6.38 ilustra la salida de un filtro de acoplamiento para las dos condiciones límite: $\theta = 0$ y $\theta = 180$ grados. Para evitar un muestreo inadecuado que surge a falta de información previa acerca de la fase θ, es razonable retener únicamente la envolvente de la salida del filtro de acoplamiento, puesto que ésta es por completo independiente del desajuste de fase θ.

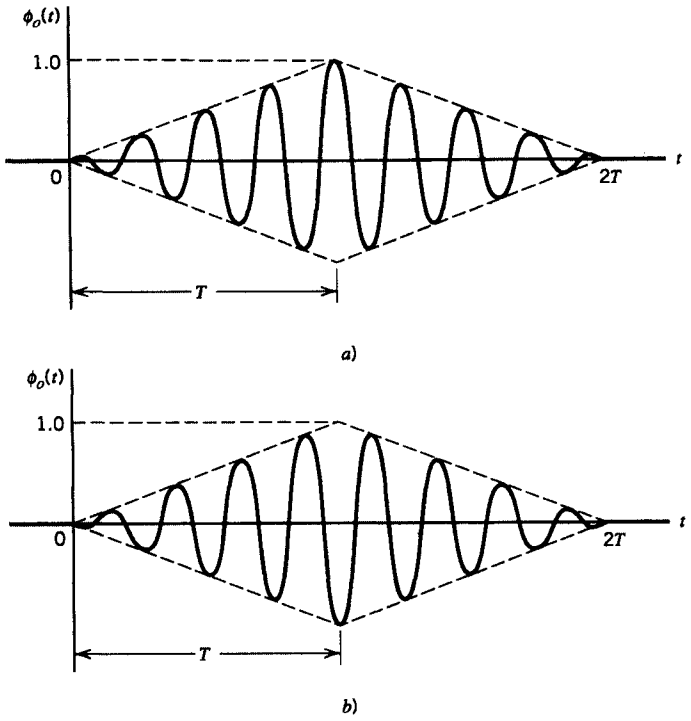


FIGURA 6.38 Salida del filtro de acoplamiento para una onda RF rectangular: *a*) $\theta = 0$ y *b*) $\theta = 180$ grados.

6.7 *Modulación ortogonal no coherente*

Con las estructuras del receptor no coherente de la figura 6.37 a nuestra disposición, podemos proceder al estudio del desempeño ante el ruido de la *modulación ortogonal no coherente* que incluye dos receptores no coherentes como casos especiales: corrimiento de frecuencia por llaveo binario no coherente y corrimiento de fase por llaveo diferencial.

Considere un esquema de transmisión de señales binarias que incluye el uso de dos señales ortogonales, $s_1(t)$ y $s_2(t)$, las cuales tienen igual energía. Durante el intervalo $0 \le t \le T$, una de estas dos señales se envía por un canal imperfecto que desfasa a la portadora una cantidad desconocida. Sean $g_1(t)$ y $g_2(t)$ las versiones recorridas de fase de $s_1(t)$ y $s_2(t)$, respectivamente. Se supone que las señales $g_1(t)$ y $g_2(t)$ se mantienen ortogonales y tienen la misma energía E, sin importar la fase desconocida de la portadora. Nos referimos a un esquema de transmisión de señales de este tipo como *modulación ortogonal no coherente*. Dependiendo de cómo definamos el par ortogonal de señales $s_1(t)$ y $s_2(t)$, la FSK y la DPSK no coherentes pueden tratarse como casos especiales de este esquema de modulación.

El canal introduce también un ruido blanco gaussiano aditivo $w(t)$ de media cero y densidad espectral de potencia $N_0/2$. En ese caso es posible expresar la señal recibida $x(t)$ como

$$x(t) = \begin{cases} g_1(t) + w(t), & s_1(t) \text{ enviada}, 0 \le t \le T \\ g_2(t) + w(t), & s_2(t) \text{ enviada}, 0 \le t \le T \end{cases} \tag{6.160}$$

El requerimiento implica utilizar $x(t)$ para discriminar entre $s_1(t)$ y $s_2(t)$, independientemente de la fase portadora.

Para este propósito empleamos el receptor que se ilustra en la figura 6.39*a*; éste consiste en un par de filtros de acoplamiento a las señales recibidas $s_1(t)$ y $s_2(t)$. Como se desconoce la fase portadora, el receptor se basa en la amplitud como el único discriminante posible. Por tanto, las salidas del filtro de acoplamiento son detectadas por la envolvente, muestreadas y después comparadas entre sí. Si la trayectoria superior en la figura 6.39*a* tiene una amplitud de salida l_1 mayor que la amplitud de salida l_2 de la trayectoria inferior, el receptor toma una decisión a favor de $s_1(t)$. Si lo inverso es correcto, la decisión es a favor de $s_2(t)$. Cuando existe igualdad, la decisión quizá se realice lanzando al aire una moneda sin alterar al azar. En cualquier caso, ocurre un error de decisión cuando el filtro de acoplamiento que rechaza la componente de señal de la señal recibida $x(t)$ tiene una amplitud de salida mayor (debido exclusivamente al ruido) que el filtro de acoplamiento por el que pasa.

De acuerdo con la explicación que se presentó en la sección 6.6, notamos que un filtro de acoplamiento no coherente (constituyendo la trayectoria superior o inferior en el receptor de la figura 6.39*a*) puede considerarse equivalente a un *receptor en cuadratura*. Este último cuenta con dos canales. Una versión del receptor en cuadratura se muestra en la figura 6.39*b*. En el canal superior, denominado el *canal en fase*, la señal recibida $x(t)$ se correlaciona con la función $\psi_i(t)$, que representa una versión escalada de la señal transmitida $s_1(t)$ o $s_2(t)$ con fase portadora

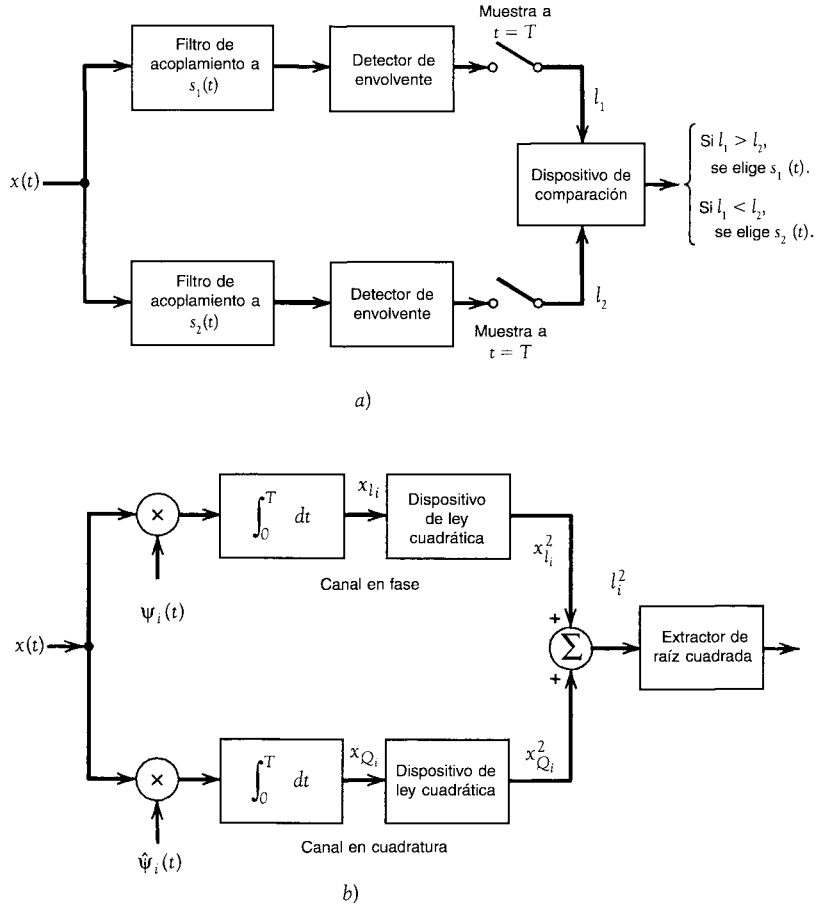


FIGURA 6.39 *a*) Receptor binario generalizado para modulación ortogonal no coherente. *b*) Receptor en cuadratura equivalente a uno de los filtros de acoplamiento en la parte *a*; el índice $i = 1, 2$.

cero. En cambio, en el canal inferior, denominado *canal en cuadratura*, $x(t)$ se correlaciona con otra función $\hat{\psi}_i(t)$, la cual representa la versión de $\psi_i(t)$ que se produce a partir del corrimiento de la fase portadora en –90 grados. De manera natural, $\psi_i(t)$ y $\hat{\psi}_i(t)$ son ortogonales entre sí.

La señal $\hat{\psi}_i(t)$, es de hecho la *transformada de Hilbert de* $\psi_i(t)$; la transformada de Hilbert se explica en el apéndice 2. Para ilustrar la naturaleza de esta relación, sea

$$\psi_i(t) = m(t)\cos(2\pi f_i t) \tag{6.161}$$

donde $m(t)$ es una señal de mensaje limitada en banda. Por lo común, la frecuencia f_i de la portadora es mayor que la componente de frecuencia más alta de $m(t)$. Entonces (de manera similar a la exposición acerca de la CAP que se presentó en la sección 6.4) la transformada de Hilbert de $\psi_i(t)$ está definida por

$$\hat{\psi}_i(t) = m(t)\,\mathrm{sen}(2\pi f_i t) \tag{6.162}$$

Puesto que

$$\cos\left(2\pi f_i t - \frac{\pi}{2}\right) = \mathrm{sen}(2\pi f_i t)$$

vemos que $\hat{\psi}_i(t)$ se obtiene en realidad de $\psi_i(t)$ desfasando la portadora $\cos(2\pi f_i t)$ en –90 grados. Una propiedad importante de la transformación de Hilbert es que una señal y su transformada de Hilbert son ortogonales entre sí. De tal manera, $\psi_i(t)$ y $\hat{\psi}_i(t)$ son ortogonales entre ellas, como ya se enunció.

La probabilidad promedio de error correspondiente al receptor no coherente de la figura 6.39*a* está dada por la fórmula simple

$$P_e = \frac{1}{2}\exp\left(-\frac{E}{2N_0}\right) \tag{6.163}$$

donde E es la energía de la señal por símbolo, y $N_0/2$ es la densidad espectral del ruido.

Deducción de la ecuación (6.163)*

Para obtener la fórmula de la ecuación (6.163) recurrimos a la equivalencia que se describe en la figura 6.39. En particular, observamos que en virtud de que se desconoce la fase portadora, el ruido a la salida de cada filtro de acoplamiento en la figura 6.39*a* tiene *dos grados de libertad*, a saber, en fase y en cuadratura. Por tanto, el receptor no coherente de la figura 6.39*a* cuenta con un total de cuatro parámetros ruidosos que son *independientes condicionalmente* dada la fase θ, así como *idénticamente distribuidos*. Estos cuatro parámetros ruidosos tienen valores de muestreo denotados por x_{I1}, x_{Q1}, x_{I2} y x_{Q2}; los primeros dos equivalen a dos grados de libertad asociados con la trayectoria superior de la figura 6.39*a*, y los últimos dos corresponden a los grados de libertad asociados con la trayectoria inferior.

El receptor de la figura 6.39*a* tiene una estructura *simétrica*. Por tanto, la probabilidad de elegir $s_2(t)$, dado que se transmitió $s_1(t)$, es la misma que la probabilidad de elegir $s_1(t)$, dado que se transmitió $s_2(t)$. Esto quiere decir que la probabilidad promedio de error puede obtenerse transmitiendo $s_1(t)$ y calculando la probabilidad de elegir $s_2(t)$, o viceversa, suponiendo que $s_1(t)$ y $s_2(t)$ son igualmente probables.

Suponga que la señal $s_1(t)$ se transmite en el intervalo $0 \le t \le T$. Ocurre un error si el ruido del canal $w(t)$ es tal que la salida l_2 de la trayectoria inferior de la figura 6.39*a* es mayor que la salida l_1 de la trayectoria superior. En ese caso el receptor toma una decisión a favor de $s_2(t)$ en

*Los lectores que no estén interesados en la derivación formal de la ecuación (6.163) pueden pasar al tratamiento del corrimiento de frecuencia por llaveo binario coherente (en la sección 6.8) y al corrimiento de fase por llaveo diferencial (en la sección 6.9), como casos especiales de modulación ortogonal no coherente, sin pérdida de continuidad.

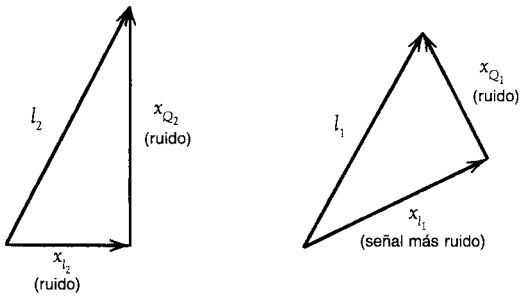


FIGURA 6.40 Interpretaciones geométricas de dos salidas de trayectoria l_1 y l_2 en el receptor no coherente generalizado.

vez de $s_1(t)$. Para calcular la probabilidad de error que se comete de esa manera, debemos tener la función de densidad de probabilidad de la variable aleatoria L_2 (representada por el valor de muestreo l_2). Puesto que el filtro en la trayectoria inferior está acoplado a $s_2(t)$, y esta última es ortogonal a la señal transmitida $s_1(t)$, se concluye que la salida de este filtro de acoplamiento se debe *exclusivamente al ruido*. Sean x_{I2} y x_{Q2} las componentes en fase y en cuadratura de la salida del filtro de acoplamiento en la trayectoria inferior de la figura 6.39*a*. Entonces, a partir de la estructura equivalente que se describe en la figura 6.39*b*, vemos que (para $i = 2$)

$$l_2 = \sqrt{x_{I2}^2 + x_{Q2}^2} \tag{6.164}$$

La figura 6.40*a* presenta una interpretación geométrica de esta relación. El ruido del canal $w(t)$ es tanto blanco (con densidad espectral de potencia $N_0/2$) como gaussiano (con media cero). De modo correspondiente, encontramos que las variables aleatorias X_{I2} y X_{Q2} (representadas por los valores de muestreo x_{I2} y x_{Q2}) son de distribución gaussiana con media cero y varianza $N_0/2$, dada la fase θ. Por tanto, podemos escribir

$$f_{X_{I2}}(x_{I2}) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{x_{I2}^2}{N_0}\right) \tag{6.165}$$

y

$$f_{X_{Q2}}(x_{Q2}) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{x_{Q2}^2}{N_0}\right) \tag{6.166}$$

En seguida utilizamos un resultado perfectamente conocido de la teoría de la probabilidad, a saber, el hecho de que la envolvente de un proceso gaussiano presenta una *distribución de Rayleigh* y es independiente de la fase θ (vea la sección 1.12). De modo específico, en la situación presente podemos establecer que la variable aleatoria L_2 [cuyo valor de muestreo l_2 se relaciona con x_{I2} y x_{Q2} mediante la ecuación (6.164)] tiene la siguiente función de densidad de probabilidad

$$f_{L_2}(l_2) = \begin{cases} \dfrac{2l_2}{N_0} \exp\left(-\dfrac{l_2^2}{N_0}\right), & l_2 \geq 0 \\ 0, & \text{en otro lado} \end{cases} \tag{6.167}$$

La figura 6.41 muestra una gráfica de esta función de densidad de probabilidad. La probabilidad condicional de que $l_2 > l_1$, dado el valor de muestreo l_1, se define por medio del área sombreada de la figura 6.41. De ese modo tenemos

$$P(l_2 > l_1 | l_1) = \int_{l_1}^{\infty} f_{L_2}(l_2)\, dl_2 \tag{6.168}$$

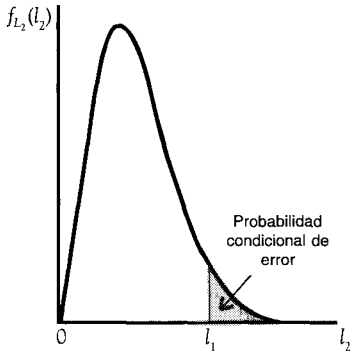


FIGURA 6.41 Cálculo de la probabilidad condicional de que $l_2 > l_1$, dada l_1.

Al sustituir la ecuación (6.167) en la (6.168) e integrar, obtenemos

$$P(l_2 > l_1 \mid l_1) = \exp\left(-\frac{l_1^2}{N_0}\right) \tag{6.169}$$

Considere a continuación la amplitud de salida l_1, relativa a la trayectoria superior en la figura 6.39*a*. Puesto que el filtro en esta trayectoria está acoplado a $s_1(t)$ y se supone que esta última se transmite, se concluye que l_1 se debe a la *señal más el ruido*. Sean x_{I1} y x_{Q1} las componentes a la salida del filtro de acoplamiento (en la trayectoria superior de la figura 6.39*a*) que están, respectivamente, en fase y en cuadratura con respecto a la señal recibida. Entonces, a partir de la estructura equivalente que se describe en la figura 6.39*b*, vemos que (para $i = 1$)

$$l_1 = \sqrt{x_{I1}^2 + x_{Q1}^2} \tag{6.170}$$

La figura 6.40*b* presenta una interpretación geométrica de esta relación. Puesto que una señal transformable de Fourier y su forma de transformada de Hilbert constituyen un par ortogonal, se llega al resultado de que x_{I1} es producto de la señal más el ruido, en tanto que x_{Q1} se origina exclusivamente del ruido. Esto significa que 1) la variable aleatoria X_{I1} representada por el valor de muestreo x_{I1} tiene una distribución gaussiana con media $\sqrt{E}$ y varianza $N_0/2$, donde E es la energía de la señal por símbolo, y 2) la variable aleatoria X_{Q1} representada por el valor de muestreo x_{Q1} es de distribución gaussiana con media cero y varianza $N_0/2$. Por tanto, es posible expresar de la siguiente manera las funciones de densidad de probabilidad de estas dos variables aleatorias independientes:

$$f_{X_{I1}}(x_{I1}) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(x_{I1} - \sqrt{E})^2}{N_0}\right) \tag{6.171}$$

y

$$f_{X_{Q1}}(x_{Q1}) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{x_{Q1}^2}{N_0}\right) \tag{6.172}$$

Puesto que las dos variables aleatorias X_{I1} y X_{Q1} son independientes, su función de densidad de probabilidad conjunta es simplemente el producto de las funciones de densidad de probabilidad dadas en las ecuaciones (6.171) y (6.172).

Para encontrar la probabilidad promedio de error, tenemos que promediar la probabilidad condicional de error dada en la ecuación (6.169) sobre todos los valores posibles de l_1. Naturalmente, este cálculo requiere el conocimiento de la función de densidad de probabilidad de las variables aleatorias L_1 representada por el valor de muestreo l_1. El método estándar implica en este caso combinar las ecuaciones (6.171) y (6.172) para determinar la función de densidad de probabilidad de L_1 resultado de la señal más el ruido. Sin embargo, esto lleva a cálculos bastante

complicados que implican el uso de las funciones de Bessel. Esta dificultad analítica quizá se haga a un lado mediante el siguiente método. Dadas x_{I1} y x_{Q1}, ocurre un error cuando, en la figura 6.39*a*, la amplitud de la salida de la trayectoria inferior l_2 debida sólo al ruido supera a l_1 que se origina de la señal más el ruido; de la ecuación (6.170) tenemos

$$l_1^2 = x_{I1}^2 + x_{Q1}^2 \tag{6.173}$$

La probabilidad de tal ocurrencia se obtiene al sustituir la ecuación (6.173) en la (6.169), lo cual se indica mediante

$$P(\text{error} \mid x_{I1}, x_{Q1}) = \exp\left(-\frac{x_{I1}^2 + x_{Q1}^2}{N_0}\right) \tag{6.174}$$

Ésta es ahora una probabilidad condicional de error, condicional sobre la salida del filtro de acoplamiento en la trayectoria superior considerando los valores X_{I1} y X_{Q1}. Esta probabilidad condicional multiplicada por la función de densidad de probabilidad conjunta de X_{I1} y X_{Q1} es entonces *la densidad de error, dadas* x_{I1} y x_{Q1}. En vista de que X_{I1} y X_{Q1} son estadísticamente independientes, su función de densidad de probabilidad conjunta es igual al producto de sus funciones de densidad de probabilidad individuales. La densidad de error resultante es una expresión complicada de x_{I1} y x_{Q1}. Sin embargo, la probabilidad de error, que es el aspecto de interés, puede obtenerse de una manera relativamente simple. Utilizamos primero las ecuaciones (6.171), (6.172) y (6.174) para evaluar la densidad de error deseada como

$$\begin{aligned} &P(\text{error} \mid x_{I1}, x_{Q1}) f_{X_{I1}}(x_{I1}) f_{X_{Q1}}(x_{Q1}) \\ &= \frac{1}{\pi N_0} \exp\left\{-\frac{1}{N_0}\left[x_{I1}^2 + x_{Q1}^2 + (x_{I1} - \sqrt{E})^2 + x_{Q1}^2\right]\right\} \end{aligned} \tag{6.175}$$

Al completar el cuadrado en el componente de la ecuación (6.175), es factible reescribir el exponente, salvo por $-1/N_0$, como

$$x_{I1}^2 + x_{Q1}^2 + (x_{I1} - \sqrt{E})^2 + x_{Q1}^2 = 2\left(x_{I1} - \frac{\sqrt{E}}{2}\right)^2 + 2x_{Q1}^2 + \frac{E}{2} \tag{6.176}$$

Ahora bien, sustituimos la ecuación (6.176) en la (6.175) e integramos la densidad de error sobre toda x_{I1} y x_{Q1}. De este modo evaluamos la probabilidad promedio de error como

$$\begin{aligned} P_e &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\text{error} \mid x_{I1}, x_{Q1}) f_{X_{I1}}(x_{I1}) f_{X_{Q1}}(x_{Q1})\, dx_{I1}\, dx_{Q1} \\ &= \frac{1}{\pi N_0} \exp\left(-\frac{E}{2N_0}\right) \int_{-\infty}^{\infty} \exp\left[-\frac{2}{N_0}\left(x_{I1} - \frac{\sqrt{E}}{2}\right)^2\right] dx_{I1} \\ &\quad \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{2x_{Q1}^2}{N_0}\right) dx_{Q1} \end{aligned} \tag{6.177}$$

Utilizamos después las siguientes dos identidades:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{2}{N_0}\left(x_{I1} - \frac{\sqrt{E}}{2}\right)^2\right] dx_{I1} = \sqrt{\frac{N_0\pi}{2}} \tag{6.178}$$

y

$$\int_{-\infty}^{\infty} \exp\left(-\frac{2x_{Q1}^2}{N_0}\right) dx_{Q1} = \sqrt{\frac{N_0\pi}{2}} \tag{6.179}$$

La identidad de la ecuación (6.178) se obtiene considerando una variable de distribución gaussiana con media $\sqrt{E/2}$ y varianza $N_0/4$, y admitiendo que el área total bajo la curva de una

función de densidad de probabilidad de una variable aleatoria es igual a la unidad; la identidad de la ecuación (6.179) resulta de un caso especial de la ecuación (6.178). De tal manera, a la luz de estas dos identidades, la ecuación (6.177) se simplifica del modo siguiente:

$$P_e = \frac{1}{2}\exp\left(-\frac{E}{2N_0}\right)$$

lo cual es el resultado deseado que se presentó antes como la ecuación (6.163).

Con esta fórmula a nuestra disposición, estamos listos para considerar la FSK binaria no coherente y la DPSK como casos especiales, lo que hacemos, respectivamente, en las siguientes dos secciones.[11]

6.8 *Corrimiento de frecuencia binaria por llaveo no coherente*

En el caso de la FSK binaria, la señal transmitida se define por medio de

$$s_i(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_i t), & 0 \le t \le T_b \\ 0, & \text{en otro lado} \end{cases} \tag{6.180}$$

donde la frecuencia f_i de la portadora es igual a uno de dos valores posibles, f_1 y f_2; para asegurar que las señales que representan estas dos frecuencias sean ortogonales, elegimos $f_i = n_i/T_b$, donde n_i es un entero. La transmisión de la frecuencia f_1 representa al símbolo 1, y la transmisión de la frecuencia f_2, al símbolo 0. En la detección no coherente de esta onda modulada en frecuencia, el receptor consiste en un par de filtros de acoplamiento seguido por detectores de envolvente, como en la figura 6.42. El filtro en la trayectoria superior del receptor se acopla a $\cos(2\pi f_1 t)$; el filtro en la trayectoria inferior, a $\cos(2\pi f_2 t)$, y en ambos casos $0 \le t \le T_b$. La salida que produce el detector de envolvente se muestrea a $t = T_b$, y se comparan sus valores. Las muestras de envolvente de las trayectorias superior e inferior en la figura 6.42 se indican como l_1 y l_2, respectivamente. Entonces, si $l_1 > l_2$, el receptor decide a favor del símbolo 1 y, en caso contrario, a favor del símbolo 0. Si $l_1 = l_2$, el receptor simplemente realiza una suposición a favor del símbolo 1 o 0.

La FSK binaria no coherente que se describe aquí es un caso especial de la modulación ortogonal no coherente con $T = T_b$ y $E = E_b$, donde T_b es la duración de bit y E_b es la energía

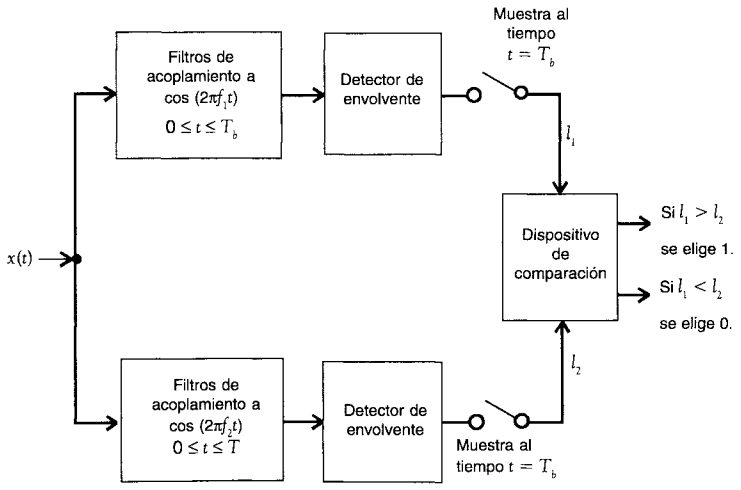


FIGURA 6.42 Receptor no coherente para la detección de señales FSK binarias.

de la señal por bit. Por tanto, utilizando la ecuación (6.163), encontramos que la *tasa de error de bits para la FSK binaria coherente* es

$$P_e = \frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right) \tag{6.181}$$

La fórmula de la ecuación (6.181) se obtiene como un caso especial de la modulación ortogonal no coherente. En el problema 6.31 consideramos el mismo aspecto utilizando un método directo que apela a la aplicación de las distribuciones de Rayleigh y de Rician; estas distribuciones pertenecen respectivamente a las variables aleatorias L_2 y L_1, cuyos valores de muestreo están definidos, en forma respectiva, por las ecuaciones (6.164) y (6.170).

6.9 *Corrimiento de fase por llaveo diferencial*

Como se subrayó en la sección 6.1, es posible considerar el *corrimiento de fase por llaveo diferencial* (DPSK) como la versión no coherente de la PSK. De ese modo se elimina la necesidad de la señal de referencia coherente en el receptor al combinar dos operaciones básicas en el transmisor: 1) *codificación diferencial* de la onda binaria de entrada y 2) *corrimiento de fase por llaveo*, de ahí el nombre de *corrimiento de fase por llaveo diferencial* (DPSK). En efecto, para enviar el símbolo 0, avanzamos la fase de la forma de onda de la señal presente en 180 grados, y para enviar el símbolo 1, dejamos sin cambio la fase de la forma de onda de la señal presente. El receptor se equipa con una función de *almacenamiento*, por lo que puede medir la *diferencia de fase relativa* entre las formas de onda recibidas durante dos intervalos de bit sucesivos. Siempre y cuando la fase desconocida θ contenida en la onda recibida varíe lentamente (es decir, lo suficientemente lento para que se considere en esencia constante a lo largo de dos intervalos de bits), la diferencia de fase entre las formas de onda recibidas en dos intervalos de bit sucesivos será independiente de θ.

La DPSK es otro ejemplo de modulación ortogonal no coherente, cuando ésta se considera sobre dos intervalos de bit. Suponga que la señal DPSK es igual a $\sqrt{E_b/2T_b}\cos(2\pi f_c t)$ para $0 \le t \le T_b$, donde T_b es la duración de bit y E_b es la energía de la señal por bit. Sea $s_1(t)$ la señal DPSK transmitida en $0 \le t \le 2T_b$ para el caso en el que tenemos el símbolo binario 1 a la entrada del transmisor en la segunda parte de este intervalo, a saber, $T_b \le t \le 2T_b$. La transmisión del símbolo 1 deja igual la fase portadora sobre el intervalo $0 \le t \le 2T_b$, y por ello definimos $s_1(t)$ como

$$s_1(t) = \begin{cases} \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & 0 \le t \le T_b \\ \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & T_b \le t \le 2T_b \end{cases} \tag{6.182}$$

Sea $s_2(t)$ la señal DPSK transmitida en $0 \le t \le 2T_b$ para el caso en el que tenemos el símbolo binario 0 en la entrada del transmisor en $T_b \le t \le 2T_b$. La transmisión del 0 avanza la fase portadora en 180 grados, por lo que definimos $s_2(t)$ como

$$s_2(t) = \begin{cases} \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t), & 0 \le t \le T_b \\ \sqrt{\dfrac{E_b}{2T_b}}\cos(2\pi f_c t + \pi), & T_b \le t \le 2T_b \end{cases} \tag{6.183}$$

Vemos de inmediato, a partir de las ecuaciones (6.182) y (6.183), que $s_1(t)$ y $s_2(t)$ son en realidad ortogonales sobre el intervalo de dos bits $0 \leq t \leq 2T_b$. En otras palabras, la DPSK es un caso especial de la modulación ortogonal no coherente con $T = 2T_b$ y $E = 2E_b$. Por tanto, utilizando la ecuación (6.163) encontramos que la *tasa de error de bits para la DPSK* está dada por

$$P_e = \frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right) \tag{6.184}$$

la cual proporciona una ganancia de 3 dB con respecto a la FSK no coherente para el mismo valor de E_b/N_0.

Generación y detección de la DPSK

El siguiente aspecto por considerar es el de la generación de las señales DPSK. El proceso de codificación diferencial en la entrada del transmisor se inicia con un primer bit arbitrario, que sirve como referencia. Sea $\{d_k\}$ la secuencia codificada diferencialmente con este bit de referencia añadido. Introducimos ahora las siguientes definiciones en la generación de estas secuencias:

- Si el símbolo binario entrante b_k es 1, el símbolo d_k queda sin cambio con respecto al bit anterior.
- Si el símbolo binario entrante b_k es 0, el símbolo d_k cambia con respecto al bit anterior.

La secuencia codificada diferencialmente $\{d_k\}$ generada de ese modo se utiliza para correr la fase de una portadora con los ángulos de fase 0 y π radianes representando, respectivamente, los símbolos 1 y 0. El proceso de codificación diferencial de fase se ilustra en la tabla 6.7. Advierta que d_k es el complemento de la suma de b_k y d_{k-1} de módulo 2.

El diagrama de bloques de un transmisor DPSK se presenta en la figura 6.43*a*. Consiste, en parte, en una red lógica y en un elemento de retardo de un bit interconectados de manera que conviertan la secuencia binaria sin procesar $\{b_k\}$ en una secuencia codificada diferencialmente $\{d_k\}$. Esta secuencia se codifica en nivel de amplitud y luego se usa para modular una onda portadora de frecuencia f_c, produciendo de ese modo la señal DPSK que se desea.

Suponga a continuación, en la detección coherente diferencialmente de la DPSK binaria, que se desconoce la fase portadora. En ese caso, al considerar que el receptor se está equipando con un canal en fase y en cuadratura, tenemos un diagrama del espacio de señales donde los puntos de la señal recibida son (A cos θ, A sen θ) y ($-A$ cos θ, $-A$ sen θ), con θ denotando la fase desconocida y A la amplitud. Esta geometría de señales posibles se ilustra en la figura 6.44. El receptor mide las coordenadas (x_{I_0}, x_{Q_0}) en el tiempo $t = T_b$ y (x_{I_1}, x_{Q_1}) en el tiempo $t = 2T_b$. La cuestión por resolver es si estos dos puntos corresponden al mismo punto de la señal o a diferentes. Admitiendo que los dos vectores $\mathbf{x}_0$ y $\mathbf{x}_1$, con puntos finales (x_{I_0}, x_{Q_0}) y (x_{I_1}, x_{Q_1}), apuntan más o menos en la misma dirección si su producto interno es positivo, podemos formular la prueba de hipótesis como sigue:

¿El producto interno $\mathbf{x}_0^T\mathbf{x}_1$ es positivo o negativo?

TABLA 6.7 *Ilustración de la generación de señal DPSK*

$\{b_k\}$		1	0	0	1	0	0	1	1
$\{d_{k-1}\}$		1	1	0	1	1	0	1	1
Secuencia $\{d_k\}$ codificada diferencialmente	1	1	0	1	1	0	1	1	1
Fase transmitida (radianes)	0	0	π	0	0	π	0	0	0

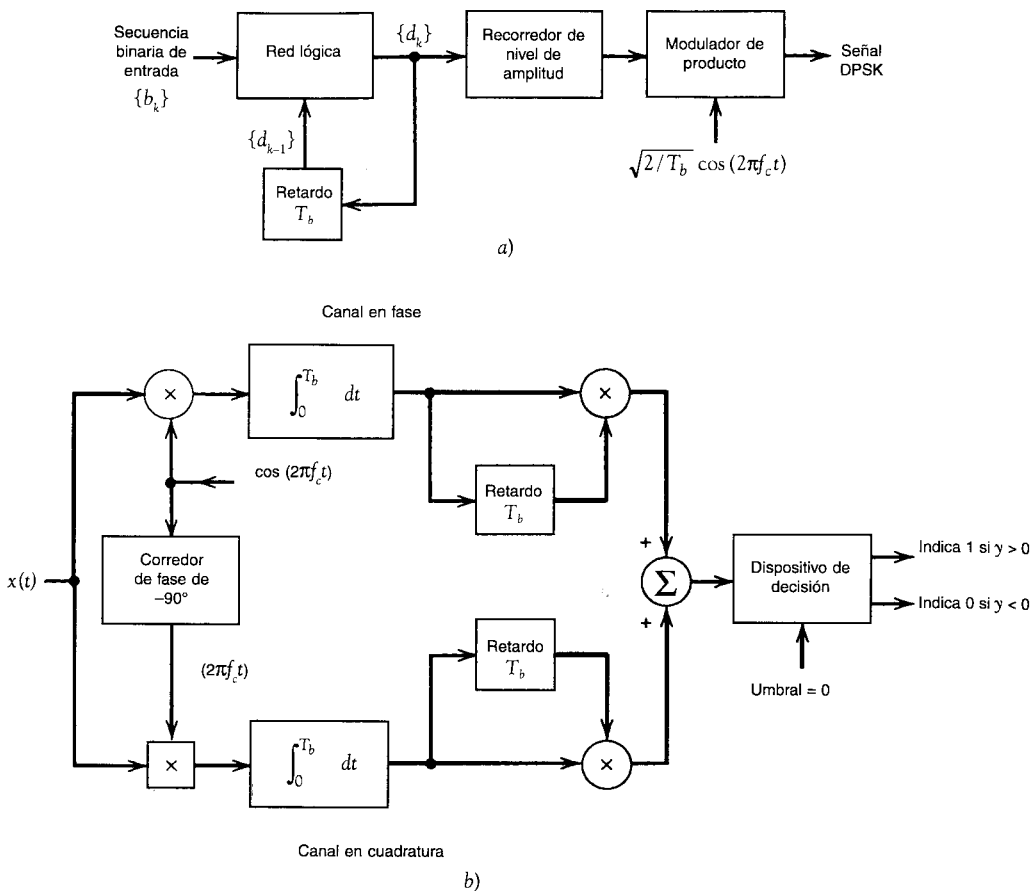


FIGURA 6.43 Diagrama de bloques de *a*) el transmisor DPSK y *b*) el receptor DPSK.

Por tanto, puede escribirse

$$x_{I_0}x_{I_1} + x_{Q_0}x_{Q_1} \underset{\text{indicado 0}}{\overset{\text{indicado 1}}{\gtrless}} 0 \tag{6.185}$$

Advertimos ahora la siguiente identidad:

$$x_{I_0}x_{I_1} + x_{Q_0}x_{Q_1} = \frac{1}{4}\left[(x_{I_0} + x_{I_1})^2 - (x_{I_0} - x_{I_1})^2 + (x_{Q_0} + x_{Q_1})^2 - (x_{Q_0} - x_{Q_1})^2\right]$$

Por tanto, al sustituir esta identidad en la ecuación (6.185) y multiplicar ambos miembros de la prueba por 4, obtenemos la prueba equivalente:

$$(x_{I_0} + x_{I_1})^2 + (x_{Q_0} + x_{Q_1})^2 - (x_{I_0} - x_{I_1})^2 - (x_{Q_0} - x_{I_1})^2 \underset{\text{indicado 0}}{\overset{\text{indicado 1}}{\gtrless}} 0 \tag{6.186}$$

Por tanto, es posible considerar al proceso de toma de decisiones como la prueba de si el punto (x_{I_0}, x_{Q_0}) está más cercano a (x_{I_1}, x_{Q_1}) o a su imagen $(-x_{I_1}, -x_{Q_1})$.

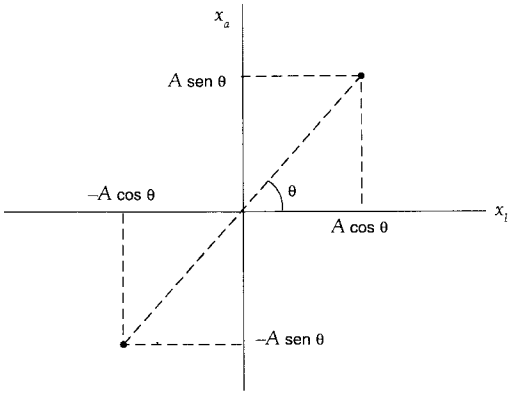


FIGURA 6.44 Diagrama del espacio de señales de la señal DPSK recibida.

De ese modo, el *receptor óptimo*[12] para la detección diferencialmente coherente de la DPSK binaria es como se muestra en la figura 6.43*b*, la cual surge directamente de la ecuación (6.185). Esta puesta en práctica requiere únicamente que se almacenen los valores *muestreados*, evitando de esa manera la necesidad de líneas de retardo ornamentales que quizá se necesiten en otros casos. La puesta en práctica equivalente del receptor que prueba elementos al cuadrado como en la ecuación (6.186) resulta más complicada, aunque su uso facilita el *análisis* en el sentido de que las dos señales por considerar son ortogonales a lo largo del intervalo $(0, 2T_b)$; por tanto, aplica el análisis de modulación ortogonal no coherente.

6.10 *Comparación de esquemas de modulación digital utilizando una sola portadora*

PROBABILIDAD DE ERROR

En la tabla 6.8 hemos resumido las expresiones para la tasa de error de bits (BER) para la PSK binaria coherente, la FSK binaria coherente convencional con decodificación de un bit, la DPSK, la FSK binaria no coherente, la QPSK coherente y la MSK coherente, cuando operan por un canal AWGN. En la figura 6.45 hemos utilizado las expresiones que se resumen en la tabla 6.8 para graficar la tasa de error de bits como función de la relación de la energía de señal por bit a la densidad espectral de ruido, E_b/N_0.

TABLA 6.8 ***Resumen de fórmulas para la tasa de error de bits de diferentes esquemas de modulación digital***

Esquema de transmisión de señales	*Tasa de error de bits*
a) PSK binaria coherente QPSK coherente MSK coherente	$\frac{1}{2}\text{ferc}\left(\sqrt{E_b/N_0}\right)$
b) FSK binaria coherente	$\frac{1}{2}\text{ferc}\left(\sqrt{E_b/2N_0}\right)$
c) DPSK	$\frac{1}{2}\exp\left(-E_b/N_0\right)$
d) FSK binaria no coherente	$\frac{1}{2}\exp\left(-E_b/2N_0\right)$

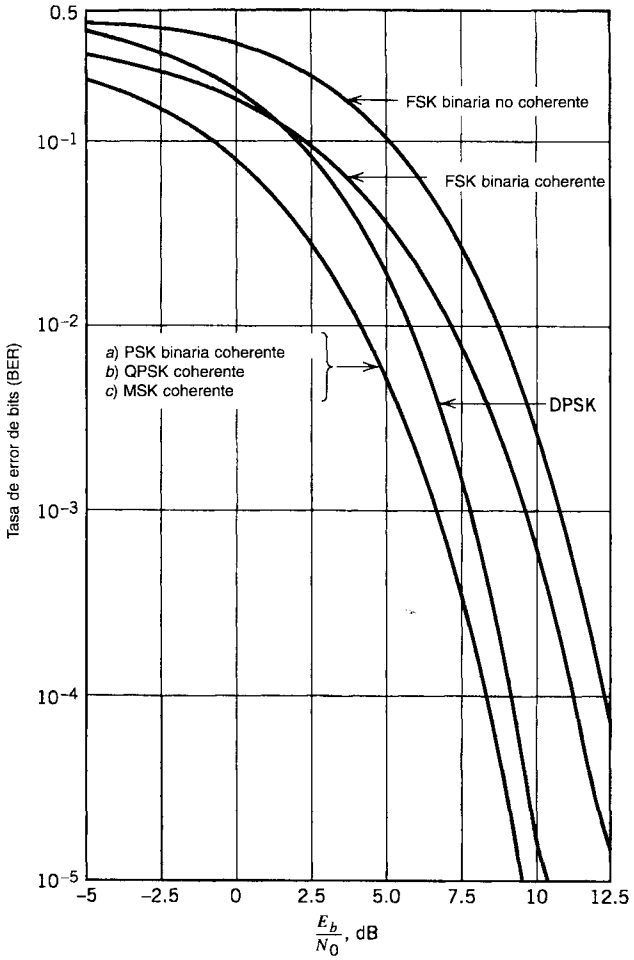


FIGURA 6.45 Comparación del desempeño ante el ruido de diferentes esquemas de PSK y FSK.

Con base en las curvas de desempeño que se muestran en la figura 6.45, el resumen de fórmulas que se da en la tabla 6.8 y las ecuaciones de definición correspondientes a los formatos de modulación pertinentes, podemos plantear los siguientes enunciados:

1. Las tasas de error de bits para todos los sistemas disminuyen monotónicamente con los valores crecientes de E_b/N_0; las curvas de definición tienen una forma similar a una *cascada*.
2. Para cualquier valor de E_b/N_0, la PSK binaria, la QPSK y la MSK coherentes producen una tasa de error de bits más pequeña que cualquiera de los demás esquemas de modulación.
3. La PSK binaria coherente y la DPSK requieren un E_b/N_0 que es 3 dB menor que los valores correspondientes para la FSK binaria coherente convencional y para la FSK binaria no coherente, respectivamente, para producir la misma tasa de error de bits.
4. A valores elevados de E_b/N_0, la DPSK y la FSK binaria no coherente se desempeñan casi tan bien (hasta dentro de 1 dB aproximadamente) que la PSK binaria coherente y la FSK binaria coherente convencional, respectivamente, para la misma tasa de bits y energía de señal por bit.
5. En la QPSK coherente se usan dos portadoras ortogonales $\sqrt{2/T}\cos(2\pi f_c t)$ y $\sqrt{2/T}\,\text{sen}(2\pi f_c t)$, donde la frecuencia de la portadora f_c es un múltiplo entero de la veloci-

dad del símbolo $1/T$, con el resultado de que los dos flujos de bits independientes pueden transmitirse de manera simultánea y detectarse de modo subsecuente en el receptor.

6. En el caso de la MSK coherente, existen dos portadoras ortogonales, a saber, $\sqrt{2/T_b}\cos(2\pi f_c t)$ y $\sqrt{2/T_b}\,\text{sen}(2\pi f_c t)$, las cuales se modulan por medio de dos pulsos de formación de símbolo antípodas $\cos(\pi t/2T_b)$ y $\text{sen}(\pi t/2T_b)$, respectivamente, a lo largo de intervalos $2T_b$, donde T_b es la duración del bit. De modo correspondiente, el receptor utiliza un proceso de decodificación de fase coherente a lo largo de dos intervalos de bit sucesivos para recuperar el flujo de bit original.
7. El esquema MSK difiere de su contraparte, la QPSK, a causa de que su receptor tiene *memoria*. En particular, el receptor MSK toma decisiones sobre dos intervalos de bits sucesivos. De esa manera, aunque la señal transmitida tiene un formato binario representado por la transmisión de dos frecuencias distintas, la secuencia de memoria en el receptor hace que asuma un diagrama del espacio de señales bidimensional. Hay cuatro puntos de mensaje, dependiendo de qué símbolo binario (0 o 1) se envió y de la historia de fase pasada de la señal FSK.

■ Eficiencia del ancho de banda de las técnicas de modulación digital M-arias

En la tabla 6.9 hemos resumido los valores característicos de los requerimientos de ancho de banda de potencia para esquemas PSK coherentes binarios y M-arios, suponiendo una probabilidad promedio del error de símbolo igual a 10^{-4} y que los sistemas operan en ambientes de ruido idénticos. Esta tabla muestra que, entre la familia de señales PSK M-arias, la QPSK (correspondiente a $M = 4$) ofrece el mejor compromiso entre los requerimientos de potencia y ancho de banda. Por esta razón, encontramos que la QPSK se utiliza bastante en la práctica. Para $M > 8$, los requerimientos de potencia se vuelven excesivos; por tanto, no se usan mucho en la práctica los esquemas PSK M-arios con $M > 8$. Además, los esquemas PSK M-arios coherentes requieren equipo considerablemente más complejo que los esquemas PSK coherentes para la generación o detección de señales, especialmente cuando $M > 8$. (La PSK 8 coherente se usa en las comunicaciones de satélite.)

Básicamente, la PSK M-aria y la QAM M-aria tienen características espectrales y de ancho de banda similares. Para $M > 4$, sin embargo, los dos esquemas presentan diferentes constelaciones de señales. En el caso de la PSK M-aria la constelación de señales es circular, en tanto que para la QAM M-aria es rectangular. Por otra parte, una comparación de estas dos constelaciones revela que la distancia entre los dos puntos de mensaje de la PSK M-aria es menor que la que hay entre los puntos de mensaje entre la QAM M-aria, para una potencia transmitida máxima fija. Esta diferencia básica entre los dos esquemas se ilustra en la figura 6.46 para $M = 16$. En consecuencia, en un canal AWGN, la QAM M-aria supera a la PSK M-aria

Tabla 6.9 ***Comparación de los requerimientos de ancho de banda entre la PSK M-aria y la PSK binaria. Probabilidad del error de símbolo = 10^{-4}***

Valor de M	$\dfrac{(\textit{Ancho de banda})_{\text{M-aria}}}{(\textit{Ancho de banda})_{\textit{binaria}}}$	$\dfrac{(\textit{Potencia promedio})_{\text{M-aria}}}{(\textit{Potencia promedio})_{\textit{binaria}}}$
4	0.5	0.34 dB
8	0.333	3.91 dB
16	0.25	8.52 dB
32	0.2	13.52 dB

De Shanmugan ([136], p. 424).

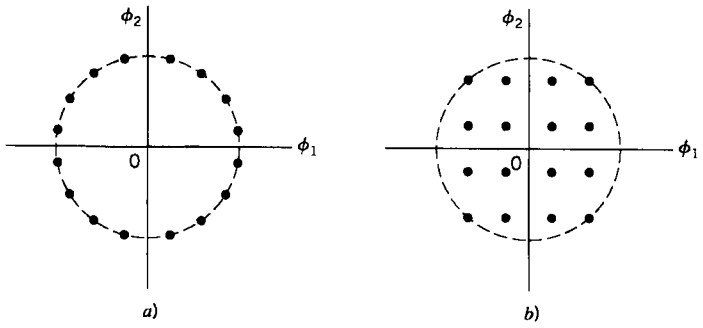


FIGURA 6.46 Constelación de señales para *a)* PSK M-aria y *b)* la QAM M-aria correspondiente, para M = 16.

correspondiente en el desempeño relativo a los errores en el caso de M > 4. Sin embargo, puede obtenerse el desempeño superior de la QAM M-aria sólo si el canal no presenta no linealidades.

Al igual que en el caso de la FSK M-aria encontramos que para una probabilidad de error fija, el aumento de M produce un menor requerimiento de potencia. Sin embargo, esta reducción de la potencia transmitida se logra a costa del incremento del ancho de banda del canal. En otras palabras, la FSK M-aria se comporta de modo opuesto a la PSK M-aria. Volveremos al tema en un contexto de la teoría de la información en el capítulo 9, por lo que integraremos un conocimiento más amplio con respecto a la comparación de los comportamientos de la PSK y la FSK M-arias.

6.11 *Módems en la banda de voz*

El "módem", contracción del término *modulador-demodulador*, es un dispositivo de conversión que facilita la transmisión y recepción de datos por la *red pública de conmutación telefónica* (PSTN).[13] Los datos de interés pueden ser señales digitales generadas por computadoras o proveedores de servicio. En este tipo de aplicaciones, la parte del modulador del módem convierte la señal digital entrante en una forma estándar adecuada para transmisión por un canal telefónico en la red telefónica mencionada. La parte del demodulador del módem recibe la salida del canal y la reconvierte en su formato original de señal digital. Incluso en otras aplicaciones, a saber, los *módems de fax* o más precisamente los módems con capacidades de facsímil, los datos quizá representen texto, gráficas, fotografías o combinaciones de ellos. En esta última aplicación, el documento de interés se codifica en una serie de elementos de imagen comprimidos (pixeles), los cuales se transmiten entonces por el canal telefónico modulando sus valores de acuerdo con un estándar de modulación predefinida. Cuando el módem de fax está en el modo de operación de recepción, la parte del demodulador del módem opera sobre la señal analógica que se recibe y descomprime la representación de datos binarios correspondientes de la señal demodulada en una réplica cercana o real de la imagen transmitida original. En seguida centraremos nuestra atención en los módems que proporcionan comunicación entre un *usuario* y un *proveedor de servicios de Internet* por la red pública de conmutación telefónica.

Por lo general, la PSTN se ha considerado como una red analógica. Sin embargo, en la actualidad se ha convertido casi por completo en una red digital. En la mayor parte de los casos, la única parte de dicha red que ha permanecido analógica (y que es probable que siga siéndolo durante los años que vienen) es el lazo local, el cual representa la conexión relativamente corta desde un lugar hasta la oficina central. De esa manera, dependiendo de cómo se usa la red pública de conmutación telefónica, podemos identificar dos clases distintas de configuraciones de módems: simétrica y asimétrica, como se describe a continuación.

■ CONFIGURACIONES DE MODEM SIMÉTRICAS

El método más simple para diseñar módems consiste en tratar la red pública de conmutación telefónica completa como una red analógica, según se indica en la figura 6.47*a*. [Recuerde del capítulo 3 que la PSTN es casi por completo digital debido al uso de la modulación por codificación de pulsos (PCM) para la transmisión de señales de voz.] En un escenario de este tipo, las conversiones analógico a digital y digital a analógico se necesitan siempre que los módems envían y reciben señales de dicha red. La configuración de módem que se describe en la figura 6.47*a* exhibe "simetría" a causa de que ambos módems son idénticos y de que la velocidad de datos *en el sentido directo* (del proveedor de servicios de Internet al usuario) es exactamente la misma que la velocidad de datos en el sentido *inverso* (del usuario al proveedor del servicio de Internet).

La configuración de módem simétrica de la figura 6.47*a* abarca un gran número de tipos de módem, que van desde velocidades de datos de 300 *b/s* hasta 36,600 b/s, como se resume en la tabla A6.7 acerca de la selección de módems estándar. El diseño de módems empieza con el corrimiento de frecuencia por llaveo, el cual provee velocidades de datos relativamente bajas. A medida que aumentó la demanda de transmisión de datos por canales telefónicos, se fueron utilizando técnicas de modulación cada vez más complicadas para mejorar el uso de la capacidad de información de dichos canales.

Considere, por ejemplo, el popular *estándar de módem V.32* que tiene las siguientes características:

Frecuencia de portadora = 1800 Hz
Velocidad de modulación = 2400 bauds
Velocidad de datos = 9600 b/s

La tasa de datos transmitidos de 9600 b/s supone una alta relación señal a ruido. El estándar V.32 especifica dos esquemas de modulación alternativos:

Codificación no redundante. Bajo este esquema, el flujo de datos entrante se divide en cuadribits (es decir, grupos de 4 bits sucesivos) y luego se transmite por el canal telefónico como QAM 16. En cada cuadribit, el dibit de entrada más significativo sufre modulación de fase, en tanto que el dibit de entrada menos significativo experimenta modulación de amplitud. Al considerar primero la modulación de fase, las consideraciones prácticas favorecen el uso de la modulación diferencial de fase, ya que el receptor sólo necesita interesarse en la detección de las cargas de fase. Este asunto se aborda utilizando un *codificador diferencial,* que consiste en una memoria sólo de lectura y en un par de unidades de retardo, como se indica en la figura 6.48*a*. Sean $Q_{1,n}Q_{2,n}$ el valor presente del dibit de entrada

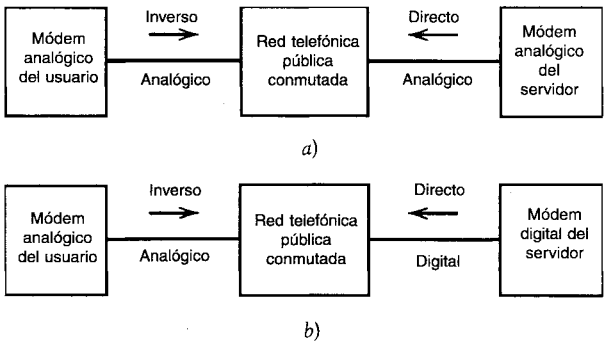


FIGURA 6.47 *a*) Panorama del ambiente de la configuración del módem simétrico: las velocidades de datos directos e inversos son iguales. *b*) Panorama ambiental de la configuración de módem "asimétrica": la velocidad de datos directos es superior que la de los inversos.

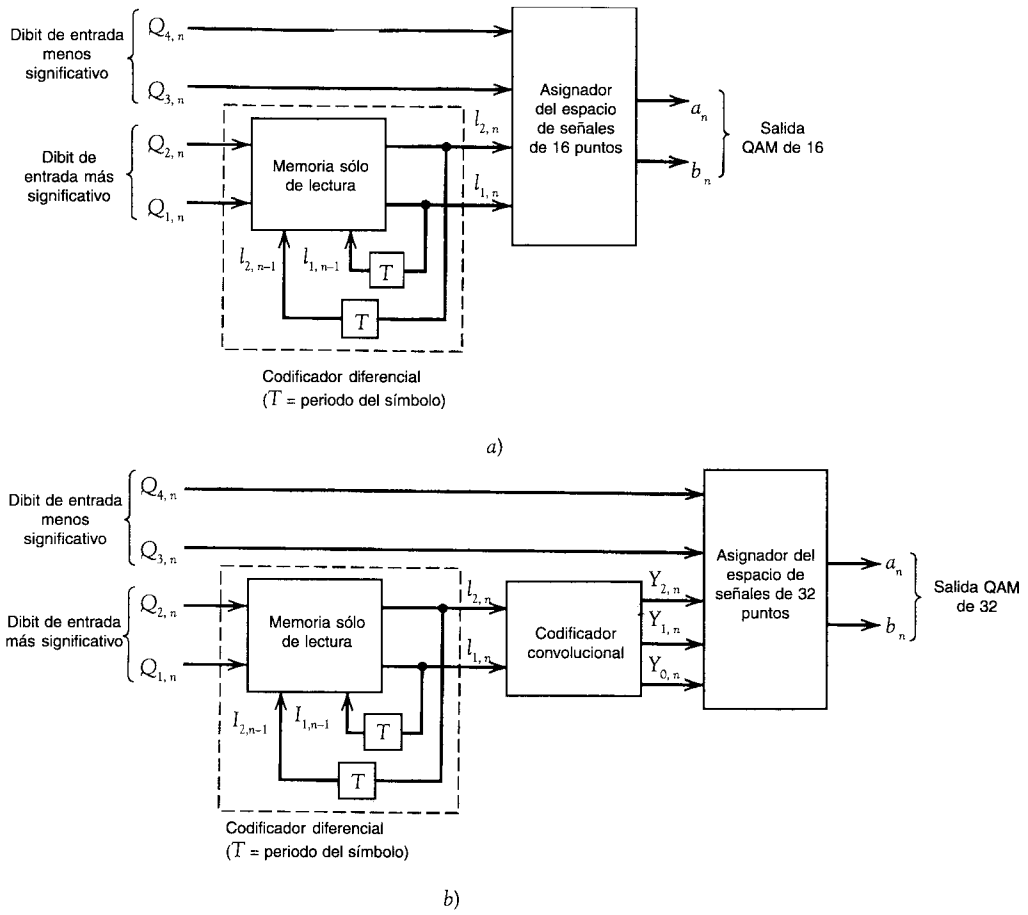


FIGURA 6.48 Diagramas de bloque del módem V.32. *a*) Codificación no redundante. *b*) Código de enramado.

más significativo, e $I_{1,n-1}I_{2,n-1}$ el valor previo de la salida de dibit correspondiente mediante el codificador. Entonces, en respuesta a los dibits $Q_{1,n}Q_{2,n}$ e $I_{1,n-1}I_{2,n-1}$, el codificador diferencial produce el dibit $I_{1,n}I_2$, el cual, a su vez, induce un cambio de fase en la señal modulada. Este cambio de fase, medido en la dirección contraria a la de las manecillas del reloj, está gobernado por el esquema de codificación Gray de la tabla 6.10. Advierta que el cambio de fase se determina por completo a partir del dibit de entrada $Q_{1,n}Q_{2,n}$. En relación con lo que se refiere a la modulación diferencial en fase, hay otro asunto que necesita tomarse en cuenta: un código para identificar los cuatro cuadrantes del espacio de señales bidimensional. Este segundo aspecto se resuelve adoptando el esquema de codificación Gray que se incluye en la figura 6.49.

Considerando ahora la modulación de amplitud, es necesario especificar un código para los cuatro valores posibles que puede asumir el dibit de entrada menos significativo, denotado por $Q_{3,n}Q_{4,n}$, digamos, en el primer cuadrante. Este aspecto se aborda adoptando el código Gray para los cuatro puntos de señal en el primer cuadrante que se muestran con el sombreado más ligero en la figura 6.49.

El aspecto final que necesita resolverse es la *invarianza rotacional de 90°*, que resulta de la codificación diferencial. Esta forma de varianza significa que la constelación QAM M-aria total se observa exactamente igual cuando se rota un múltiplo entero de 90°, sin

TABLA 6.10 ***Cambios de fase inducidos por la codificación diferencial en el módem V.32 debidos a dibits de entrada variables***

Dibit de entrada presente		*Cambio de fase (grados)*
$Q_{1,n}$	$Q_{2,n}$	
0	0	90
0	1	0
1	0	180
1	1	270

que importe si está codificada o decodificada; en ese caso el receptor puede decodificar correctamente la secuencia de mensaje transmitida cuando la fase del oscilador local difiere de la fase portadora en un múltiplo entero de 90°. Este requerimiento final se satisface completando los códigos Gray para los puntos de mensaje en los tres cuadrantes restantes de la manera que se señala en la figura 6.49. Las líneas punteadas se incluyen en la figura 6.49 para ilustrar la invarianza rotacional de 90 grados.

Reuniendo todos estos aspectos en la modulación de amplitud y en fase combinada, obtenemos la constelación QAM 16 que se mostró antes en la figura 6.17*a*, la cual se reproduce aquí como la figura 6.50*a*. En forma correspondiente, el sistema de codificación consiste en un codificador diferencial seguido por un asignador del espacio de señales de 16 puntos, como se indica en la figura 6.48*a*. El módem V.32 configurado de esa manera se dice que es no redundante debido a que, con 16 puntos de constelación, la palabra de código transmitida de 4 bits *no* tiene bits redundantes.

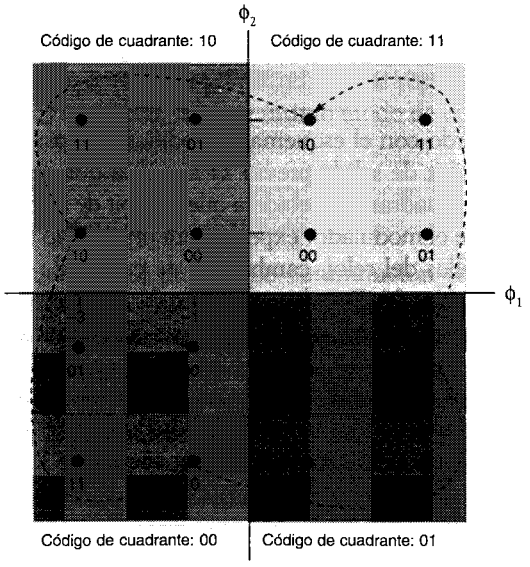


FIGURA 6.49 Ilustración de la codificación Gray de los cuatro cuadrantes y dibits en cada cuadrante para el módem V.32. Las flechas punteadas ilustran la invarianza rotacional de 90°.

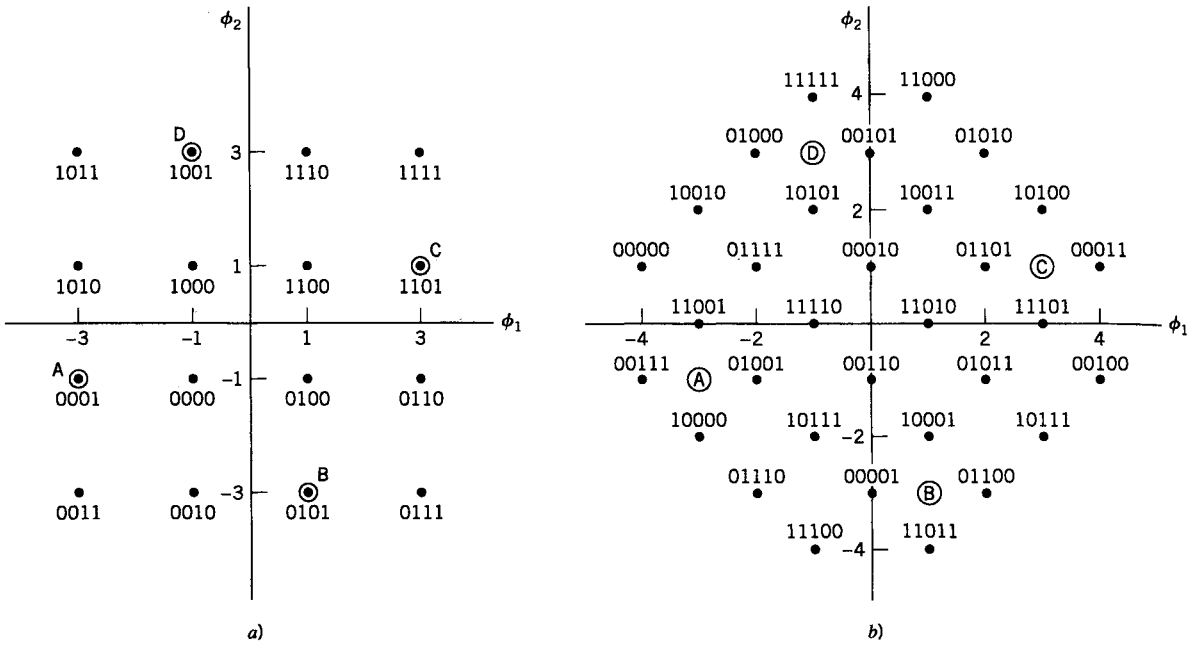


FIGURA 6.50 *a*) Constelación de señales del módem V.32 utilizando codificación no redundante. *b*) Constelación de señales del módem V.32 utilizando codificación de enramado.

Como un ejemplo ilustrativo de cómo opera este módem V.32 particular, considere que el grupo presente de 4 bits de entrada es 1001 y que el dibit producido a salida por el módem es igual a 11. Para este ejemplo, tenemos entonces

$$Q_{1,n}Q_{2,n} = 10$$
$$Q_{3,n}Q_{4,n} = 01$$
$$I_{1,n-1}I_{2,n-1} = 11$$

Entonces, de acuerdo con el esquema de codificación para los cuatro cuadrantes especificado en la figura 6.49, el dibit de salida previo 11 significa que el modulador estaba residiendo anteriormente en el primer cuadrante. Debido a que el dibit de entrada correspondiente es 10, se concluye de la tabla 6.10 que el modulador experimenta un cambio de fase de 180° en la dirección contraria a la de las manecillas del reloj, cambiando de ese modo su operación a la del tercer cuadrante identificado por el dibit 00. Por último, siendo 01 el valor presente del dibit $Q_{3,n}Q_{4,n}$, menos significativo, el modulador produce una señal cuyas coordenadas son $a_n = -3$ (a lo largo del eje ϕ_1) y $b_n = -1$ (a lo largo del eje ϕ_2). Esta salida corresponde a la palabra de código 0001.

Cuando la relación señal a ruido no es suficientemente alta, el módem V.32 cambia a su modo QPSK, operando a una velocidad reducida de 4800 b/s. En este modo de operación, los cuatro estados del modo se representan por medio de los puntos marcados A, B, C y D en la figura 6.50*a*.

Codificación de enramado

La codificación de enramado es un esquema de corrección de errores directo donde la codificación y la modulación se consideran como una entidad combinada en vez de dos operaciones

independientes. La figura 6.48*b* presenta el sistema de codificación del módem V.32 con codificación de enramado. La secuencia de datos entrante se divide en cuadribits, aunque a diferencia del caso de la codificación no redundante, éstos se transmiten por el canal como una señal QAM 32.

Como se indica en la figura 6.48*b*, el codificador de enramado implica el uso de un *codificador convolucional* que opera sobre la salida del codificador diferencial. (Los codificadores convolucionales se explican en el capítulo 10.) Sin embargo, la elección de la codificación convolucional se restringe en el módem V.32 para adaptarse al uso de la codificación diferencial (es decir, invarianza rotacional de 90°). En realidad, este requerimiento no puede satisfacerse por medio de un codificador convolucional. En cambio, este último debe ser *no lineal*;[14] vea el problema 10.30.

El proceso de codificación de datos en el módem V.32 con codificación de enramado se ejecuta en tres etapas:

1. El codificador diferencial en la figura 6.48*b*, en respuesta al dibit de entrada presente $Q_{1,n}Q_{2,n}$ y el dibit previo codificado diferencialmente, $I_{1,n-1}I_{2,n-1}$, producen el dibit $I_{1,n}I_{2,n}$.
2. El dibit presente codificado diferencialmente $I_{1,n}I_{2,n}$ es la entrada para el codificador convolucional en la figura 6.48*b*, el cual produce una salida de 3 bits. Uno de ellos es un *bit de verificación de paridad* denotado por $Y_{0,n}$. El valor de $Y_{0,n}$ depende de los otros dos bits, $Y_{1,n}$ y $Y_{2,n}$, producidos por el codificador convolucional.
3. Los bits $Y_{0,n}$, $Y_{1,n}$ y $Y_{2,n}$ producidos por el codificador convolucional, junto con el dibit de entrada menos significativo $Q_{3,n}Q_{4,n}$ se aplican al asignador del espacio de señales en la figura 6.48*b*, el cual elige uno de los estados en la constelación de 32 puntos que se muestra en la figura 6.50*b* como la salida del módem.

El bit de verificación de paridad $Y_{0,n}$ ofrece un módem con mejor inmunidad del código de enramado a los deterioros del canal que un módem V.32 con codificación no redundante, ventaja que se obtiene sin un incremento en los requerimientos del ancho de banda. En términos cuantitativos, la codificación de enramado proporciona una ganancia de codificación efectiva de 4 dB en comparación con la QAM 16. La *ganancia de codificación* expresa qué tanta más energía de la señal por bit de datos necesita un módem no codificado para el mismo nivel de desempeño ante el ruido.

Sin embargo, para que se aproveche esta ventaja de la codificación de enramado en la práctica, la relación de señal a ruido debe ser lo suficientemente alta. En otro caso, el módem V.32 se conmuta a su modo de operación QPSK, el cual se representa por cuatro estados denominados A, B, C y D en la figura 6.50*b*. En este último modo de operación, la velocidad de datos del módem se reduce hasta 4800 b/s.

■ Configuraciones de módem asimétricas

Para un uso más eficiente de la red pública de conmutación telefónica, debemos tratar ésta como lo que realmente es: una *red enteramente digital que no es lineal*. En particular, ya que los proveedores de servicios de Internet se implantan digitalmente, se elimina la necesidad de conversión analógica-digital en el módem ISP, lo cual significa que la comunicación entre la ISP y la PSTN puede ser totalmente digital, como se muestra en la figura 6.47*b*. Sin embargo, el módem del usuario tiene que permanecer analógico debido a que el lazo local es analógico, por lo que requiere el uso de conversiones analógica-digital y digital-analógica cada vez que el módem del usuario envía señales a y las recibe de la red pública de conmutación telefónica. La configuración de módem que se describe en la figura 6.47*b* es "asimétrica", de manera que es posible que la velocidad de los datos de transmisión en sentido directo sea mucho mayor que la correspondiente a la velocidad de datos de transmisión en sentido inverso, como se explica a continuación.

Como se mencionó antes, una PSTN digital se basa en el uso de la PCM para la transmisión de señales de voz. Las características importantes del sistema para la explicación presente son las siguientes (vea el capítulo 3):

- Velocidad de señalización de datos de 64 kb/s, la cual se conforma por una velocidad de muestreo de 8 kHz y la representación de cada muestra de voz mediante una palabra de código de 8 bits.
- Ley de compresión y expansión de 15 segmentos (por ejemplo, una ley μ logarítmica con $\mu = 255$) para comprimir la señal de voz en el transmisor y expandirla en el receptor.

De la explicación acerca de la PCM que se presentó en el capítulo 3 recordamos también que la cuantización sólo afecta a la conversión analógica digital, pero *no* a la conversión digital-analógica. Esta observación tiene un efecto profundo sobre la estrategia óptima para el diseño de los módems asimétricos.

Suponga que no hay conversión analógica-digital entre un módem digital en el ISP y la porción digital de la PSTN, y que el transmisor digital conectado directamente del módem está diseñado para usar en forma apropiada los 256 niveles de umbral (discretos) espaciados no uniformemente de la PSTN digital. En ese caso, debido a que la conversión digital-analógica no sufre en absoluto ningún efecto del ruido de cuantización, se concluye que la información transmitida por el módem digital del ISP llega sin pérdida de ninguna clase al módem analógico del usuario. Con base en estos argumentos debe ser posible, en teoría, transmitir datos del ISP al usuario a una velocidad igual a la velocidad de datos de 64 kb/s de la PSTN digital. Sin embargo, las limitaciones de sistema inherentes a esta última reducen la velocidad de datos que puede conseguirse por debajo de 56 kb/s, como se explica más adelante.

Módem digital

A partir de la descripción de un canal de banda de voz PCM que se presenta en el capítulo 3, encontramos que el diseño del módem digital está restringido por tres factores *fuera* de nuestro control. Las restricciones de diseño son:

1. Una velocidad de muestreo $f_s = 8$ kHz.
2. Un conjunto de $M = 256$ niveles de umbral permisibles incorporados en la construcción del compresor (es decir, la porción de transmisor del compresor-ensanchador).
3. Un filtro de banda base (antitraslape) de aproximadamente 3.5 kHz de ancho de banda, incorporado en el extremo frontal del transmisor PCM.

De acuerdo con estas restricciones, podemos establecer los criterios que fundamentan el diseño del módem digital del modo siguiente:

> Diséñese una señal $s(t)$ a la entrada del módem digital de tal manera que cada una de sus muestras tomadas a una frecuencia $f_s = 8$ kHz corresponda con uno de los $M = 256$ niveles de umbral del compresor, y la señal transmitida satisfaga el criterio de Nyquist para interferencia de intersímbolo cero.

(El criterio de Nyquist para interferencia de intersímbolo cero se explica en el capítulo 4.)

Una realización del módem digital

Una solución de este problema de diseño de señal se vuelve en particular difícil por el hecho de que el filtro del transmisor PCM tiene un ancho de banda de casi 3.5 kHz y no de 4 kHz (la mitad de la frecuencia de muestreo f_s). La aplicación inmediata de esta restricción es que en lugar del conjunto deseado de 8000 muestras, sólo podemos generar $2 \times 3500 = 7000$ muestras *independientes* cada segundo de acuerdo con el criterio de Nyquist para interferencia

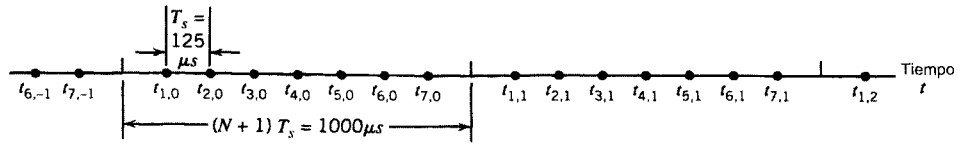


FIGURA 6.51 Grupo de N muestras uniformemente espaciadas, repitiéndose cada $(N + 1)T_s$ segundos.

de intersímbolo. ¿Cómo ajustaremos entonces 7000 muestras independientes por segundo dentro del marco preestablecido de 8000 muestras por segundo?

Para dar respuesta a esta pregunta fundamental, utilizamos la forma *equivalente no uniforme recurrente* del teorema de muestreo. Para ser más específicos, considere la situación que se describe en la figura 6.51, donde las muestras se dividen en grupos, cada uno de ellos conteniendo N muestras espaciadas uniformemente, y con los grupos teniendo un periodo recurrente de $(N + 1)T_s$ segundos, donde $T_s = 1/f_s$. La ilustración que se presenta en la figura 6.51 corresponde al problema presente: $T_s = 125$ μs y $N = 7$. Los instantes de muestreo en la distribución no uniforme de la figura 6.51 se escriben como

$$\begin{aligned} t_{k,l} &= t_k + (N+1)lT_s \qquad k = 1, 2, \ldots, N \\ &= (k-1)T_s + (N+1)lT_s, \qquad l = 0, \pm 1, \pm 2, \ldots \end{aligned} \tag{6.187}$$

Todo está listo ahora para que definamos la señal limitada en banda $s(t)$ de la manera siguiente:[15]

$$s(t) = \sum_{l=-\infty}^{\infty} \sum_{k=1}^{\infty} s(t_{k,l})\,\psi_k(t - (N+1)lT_s) \tag{6.188}$$

donde la *función de interpolación* $\psi_k(t)$ está definida por

$$\psi_k(t) = \operatorname{senc}\left(\frac{t - t_k}{(N+1)T_s}\right) \prod_{\substack{q=1 \\ q \neq k}}^{N} \frac{\operatorname{sen}\left(\dfrac{\pi}{(N+1)T_s}(t - t_q)\right)}{\operatorname{sen}\left(\dfrac{\pi}{(N+1)T_s}(t_k - t_q)\right)} \tag{6.189}$$

Al calcular la ecuación (6.189) para $N = 7$, obtenemos los siete *pulsos estándar* que se grafican en la figura 6.52, donde el tiempo está normalizado con respecto al periodo de muestreo T_s. Estos pulsos exhiben las siguientes propiedades:

- Cada pulso estándar está normalizado de manera que tenemos

$$\psi_k\left(\frac{t_k}{T_s}\right) = \psi_k(k-1) = 1 \qquad \text{para } k = 1, 2, \ldots, 7$$

  Sin embargo, advierta que el pico del k-ésimo pulso *no* ocurre en el tiempo $t_k = (k - 1)T_s$.
- Para $k = 1, 2, \ldots, 7$, el pulso $\psi_k(t/T_s)$ se hace cero en los tiempos $t \neq (k - 1)T_s$ módulo $(N + 1)$, salvo en aquellos tiempos que son congruentes con $t = (-1)$ módulo $(N + 1)$.

Por tanto, el esquema de transmisión de señales para el módem digital consiste en un esquema de modulación de amplitud de pulsos no uniformes recurrentes. Las amplitudes de 7 muestras espaciadas uniformemente en cada grupo de 8 muestras se determinan mediante la secuencia de datos entrante y en conformidad con los niveles de umbral del compresor en el transmisor PCM. En efecto, estas 7 muestras constituyen muestras independientes que son

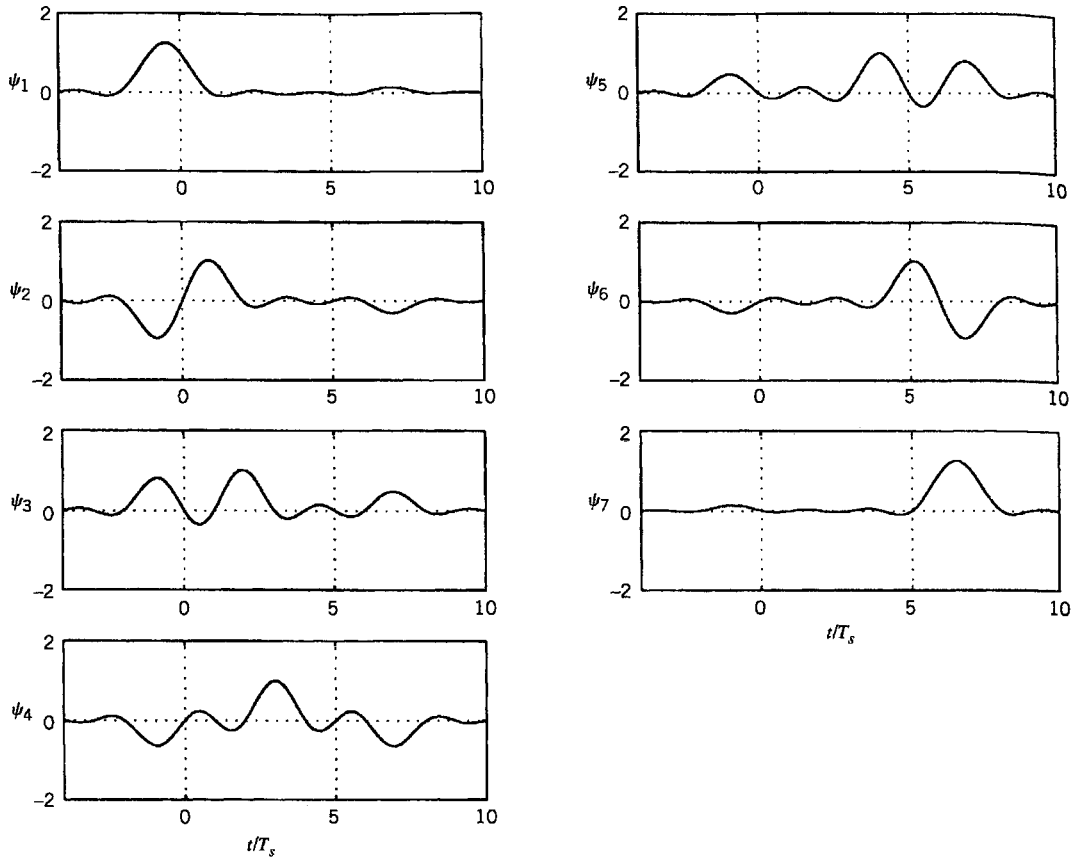


FIGURA 6.52 Formas de onda de un módem digital de los pulsos estándar $\psi_k(t)$, $k = 1, 2, ..., 7$.

responsables de llevar la secuencia de datos entrante a través de la PSTN cada 1000 μs. Además, entregan los datos al receptor con interferencia de intersímbolo *cero*. Las "ocho" muestras restantes están determinadas por completo por las muestras independientes y las conoce de antemano el sistema. No contienen información y, por tanto, se descartan en el receptor. De esta manera, el módem digital es capaz de transmitir datos digitales a través de la PSTN casi sin errores a una velocidad igual a 56 kb/s, la cual se calcula como sigue:

$$7 \times 1000 \times \log_2 256 = 56{,}000 \text{ b/s}$$

Conviene un último comentario. El pulso estándar $\psi_k(t)$ puede construirse de modo que decaiga a una tasa más rápida que $1/t$. Al hacerlo de ese modo, simplemente sustituimos la función senc en la ecuación (6.189) por un pulso de Nyquist con una caída de una manera similar a la que se describe en el capítulo 4.

Otra realización del módem digital

El tipo de módem digital que se acaba de describir es *bidireccional*, suponiendo que ambos extremos del enlace de datos son analógicos. Sin embargo, una solución más simple al problema de diseño del módem digital resulta cuando un extremo del enlace es digital y son posibles velocidades de datos asimétricas.[16]

Considere lo que sucede cuando una secuencia de datos consistente en *objetos* (es decir, palabras de 8 bits) llega a la PSTN. Ahí serán tratados como objetos que representan habla codificada de acuerdo con la ley μ o la ley A, dependiendo de la parte de la palabra donde se

ubique la PSTN. En consecuencia, el convertidor D/A que acciona al módem analógico produce una señal en tiempo continuo definida por

$$s(t) = \sum_k a(c_k) g(t - kT_s) \tag{6.190}$$

donde c_k es el k-ésimo objeto en la secuencia de datos, $a(c_k)$ es el nivel de representación especificado por la ley de compresión-expansión pertinente, T_s es el intervalo de muestreo (igual a 125 μs), y $g(t)$ es una función de interpolación limitada en banda a una frecuencia por debajo de $1/2T_s$, o aproximadamente 4 kHz, para satisfacer la parte de la reconstrucción del teorema de muestreo; vea la sección 3.2.

En la operación normal de la PSTN, la señal $s(t)$ representa una señal del habla reconstruida. Sin embargo, en el caso de datos de entrada, $s(t)$ se asemeja al ruido. En cualquier caso, desde una perspectiva teórica de la comunicación, la señal $s(t)$ en la ecuación (6.190) puede considerarse como una señal *modulada por amplitud de pulsos*. Aquí se encuentra la base teórica para el diseño del módem digital. Específicamente, el diseño se basa en una constelación de señales como en un módem analógico, salvo que esta última se construye a partir de *símbolos PCM unidimensionales* en vez de símbolos QAM bidimensionales.

De modo ordinario, la velocidad de datos que puede lograrse mediante un módem digital está limitada a cerca de 56 kb/s debido a los siguientes factores:

1. Los niveles internos del compresor-expansor en la PSTN están espaciados muy estrechamente, como se indica en la tabla 3.4; por tanto, están propensos a la interferencia de intersímbolo residual y al ruido que sigue del ecualizador del módem.
2. Los bits menos significativos son extraídos de la secuencia de datos para diversos fines internos de la PSTN; esta "extracción de bits" puede ser tan grande como 8 kb/s (aunque suele ser menor que esta cifra) y siempre en un patrón periódico.

Módem analógico

A diferencia del módem digital, el desempeño con respecto al ruido del módem analógico está limitado en esencia por el ruido de cuantización en la ley μ o la ley A que gobierna la operación del compresor-expansor PCM. Por lo común, la relación señal a ruido en un buen canal de banda de voz PCM es del orden de 34 a 38 dB. El otro deterioro del canal que limita la operación del módem analógico es el efecto de la limitación en banda que imponen los filtros de antitraslape e interpolación, el cual, como ya se mencionó, casi siempre es cercano a 3.5 kHz.

Una elección compleja para el módem analógico es el estándar de *módem V.34*, que opera a velocidades que se extienden hasta 33.6 kb/s. El criterio de diseño fundamental de este módem incorpora cinco características distintivas.[17]

1. *Superconstelación QAM 960.*
La constelación de señales que dice que será super o anidada en el sentido de que está compuesta por cuatro constelaciones: la constelación QAM que se muestra en la figura 6.53 con 240 puntos de mensaje, y sus versiones rotadas en 90, 180 y 270 grados.

2. *Ancho de banda adaptable.*
El transmisor confirma el canal enviando un conjunto de tonos, que permiten medir la relación de señal a ruido a la salida del canal como una función de la frecuencia. De esta manera se permite al módem elegir la frecuencia y el ancho de banda de portadora apropiados de acuerdo con los resultados de la prueba y las velocidades de símbolo disponibles.

3. *Velocidades de bit adaptables.*
Durante el entrenamiento del receptor, la velocidad de bit se elige de acuerdo con la estimación del receptor de la máxima velocidad de bit, la cual puede soportar el módem a velocidades de error de bit tan bajas como de 10^{-6} a 10^{-5}.

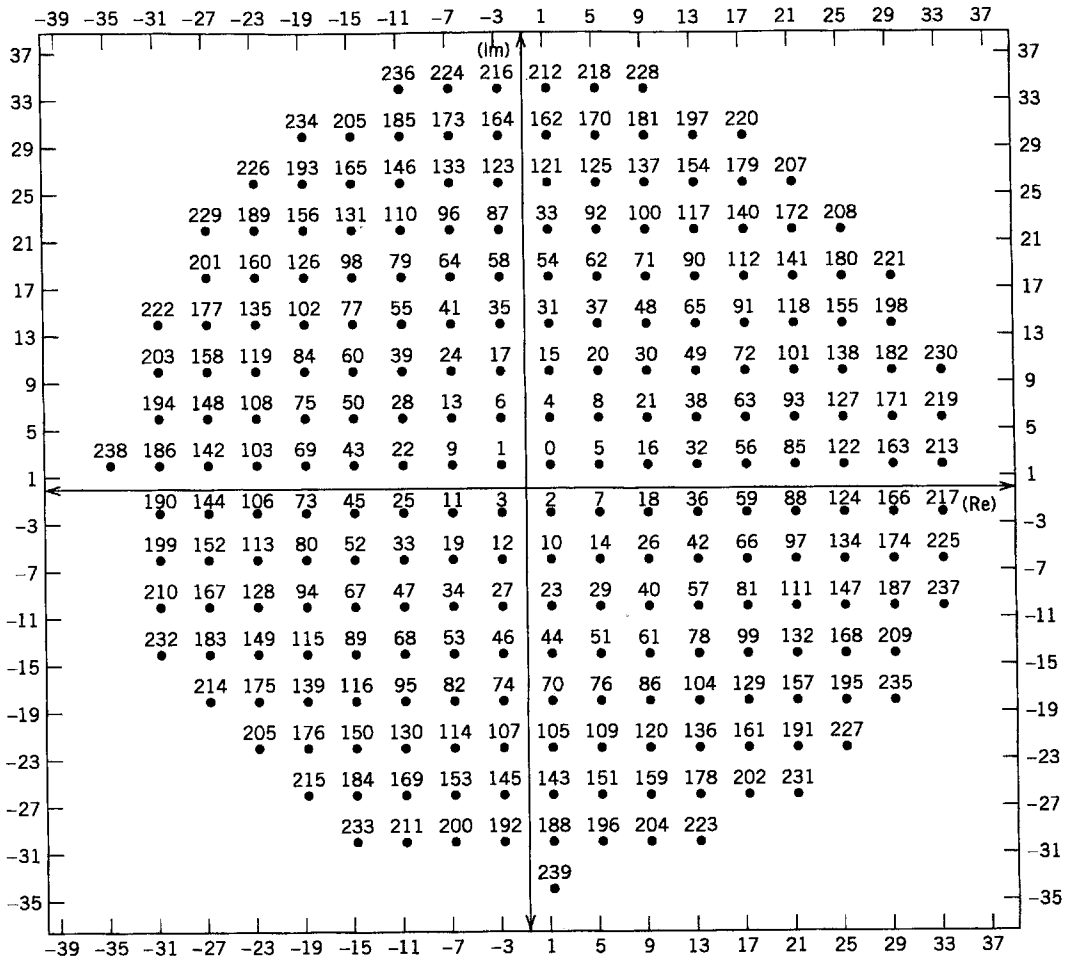


FIGURA 6.53 Superconstelación de cuatro del módem V.34 con 240 puntos de señales. La superconstelación completa se obtiene combinando las versiones rotadas de estos puntos en 0, 90, 180 y 270 grados. (Tomada de Forney y otros con autorización del IEEE.)

4. *Codificación de enramado.*
Esta técnica de codificación de control de errores se utiliza para ofrecer una ganancia de codificación efectiva de casi 3.6 dB; existe un código de enramado opcional más poderoso con una ganancia de codificación efectiva de aproximadamente 4.7 dB.

5. *Ecualización de decisión realimentada.*
Para hacer un uso total de un ancho de banda disponible del canal telefónico, incluso en frecuencias cercanas a los bordes de la banda donde la atenuación puede ser tan grande como entre 10 y 20 dB, se utiliza un ecualizador de decisión realimentado (DFE). (El DFE se explica en el capítulo 4.) Sin embargo, no es una cuestión inmediata combinar la codificación con el DFE debido a que la realimentación requiere decisiones inmediatas, en tanto que la codificación implica inherentemente el retardo de la decodificación. Para superar este problema, la sección de realimentación del DFE se mueve al transmisor, lo que es posible mediante el uso de la precodificación de Tomlinson-Harashima. (Esta forma de ecualización por medio de la precodificación se explica de manera breve en la nota 12 del capítulo 4.)

Módem V.90

El estándar del módem V.90 abarca módems digitales y analógicos. El módem digital en el extremo ISP se basa en la segunda realización que se describió antes; éste envía datos en forma directa a razón de 56 kb/s. El módem analógico en el extremo del usuario es un estándar de módem V.34, que transmite datos en sentido inverso a razón de 33.6 kb/s. Estas dos velocidades altamente diferentes confirman la naturaleza asimétrica del módem V.90.

La característica sobresaliente del módem V.90, a saber, la velocidad de datos en sentido directo de 56 kb/s, lo hace adecuado para su uso en Internet para bajar gráficas en páginas web intensivas, audio y video a velocidades cercanas a las de la ISDN.

6.12 *Modulación multicanal*

La *línea digital asimétrica de suscriptor* (*ADSL*), que se describió en la sección 4.8, es un sistema de transmisión de datos capaz de ejecutar velocidades de megabits sobre líneas telefónicas de par trenzado existentes. De manera específica, la ADSL opera a una velocidad de datos en sentido directo de hasta 9 Mb/s y una velocidad de datos en sentido inverso de hasta 1 Mb/s. Estas velocidades de transmisión de señales de datos encajan perfectamente en los requerimientos de acceso de Internet. (Como se mencionó en la sección 4.8, la velocidad de bits en sentido inverso debe ser alrededor de 10% de la de sentido directo para la operación eficiente del protocolo de Internet.) El desafío en el diseño de la ADSL radica en desarrollar un código de línea que explote la velocidad de información del canal al mayor grado posible. La modulación de amplitud/fase sin portadora (CAP), que se explicó en la sección 6.4, proporciona un método para resolver este difícil problema de la transmisión de datos. Otro método es usar una técnica de modulación igualmente elegante denominada multitonos discretos. Este último método es una forma de *modulación multicanal*[18] que permite que la característica del modulador sea una función de la característica del canal medido. Por ello, empezamos en esta sección con la modulación multicanal y continuamos en la siguiente con el tema de multitonos discretos.

La idea básica de la *modulación multicanal* se fundamenta en un principio que suele usarse en ingeniería: *divide* y *vencerás*. De acuerdo con este principio, un problema difícil se resuelve dividiéndolo en varios problemas más simples y combinando después esas soluciones. En el contexto de nuestra explicación presente, el problema difícil es el de la transmisión de datos por un canal de banda ancha con severa interferencia de intersímbolo, y los problemas más simples se ejemplifican mediante la transmisión de datos por canales AWGN. Es posible resumir entonces la esencia de la modulación multicanal en la forma siguiente:

> La transmisión de datos por un canal difícil se transforma mediante el uso de técnicas avanzadas de procesamiento de señales en la transmisión en paralelo del flujo de datos dados por un número mayor de subcanales, de tal manera que cada subcanal pueda considerarse efectivamente como un canal AWGN.

Naturalmente, la velocidad de datos total es la suma de las velocidades de datos individuales sobre los subcanales que operan en paralelo.

Capacidad del canal AWGN

Con base en el material de antecedentes y preliminares que se presenta en el capítulo inicial recordamos que, de acuerdo con el *teorema de la capacidad de información de Shannon*, la capacidad de un canal AWGN (sin interferencia de intersímbolo) está definida por

$$C = B \log_2(1 + \text{SNR}) \text{ b/s} \tag{6.191}$$

donde B es el ancho de banda del canal y SNR denota la relación señal a ruido medida a la salida del canal. Una prueba de este importante teorema se presenta de manera formal en el capítulo 9. Por ahora basta decir que para una SNR dada, podemos transmitir datos por un canal AWGN de ancho de banda B a una velocidad máxima de C bits por segundo con una probabilidad de error arbitrariamente pequeña, siempre y cuando utilicemos un sistema de codificación de complejidad suficientemente alta. De manera equivalente, es posible expresar la capacidad C en bits por transmisión o uso de canal como

$$C = \frac{1}{2}\log_2(1 + \text{SNR}) \quad \text{bits/transmisión} \tag{6.192}$$

En la práctica, casi siempre encontramos que un sistema de codificación físicamente realizable debe transmitir datos a una velocidad R menor que la velocidad máxima posible C para la cual será confiable. Para un sistema que pueda ponerse en práctica y que opere a una probabilidad del error de símbolo suficientemente baja, necesitamos consecuentemente introducir una *brecha de relación señal a ruido* o sólo *brecha* denotada por símbolo Γ. La brecha es una función de la probabilidad previsible del error de símbolo P_e y el sistema de codificación de interés. Proporciona una medida de la "eficiencia" de un sistema de codificación con respecto al sistema de transmisión ideal de la ecuación (6.192). Con C denotando la capacidad del sistema de codificación ideal y R representando la capacidad del sistema de codificación correspondiente que puede ponerse en práctica, la brecha se define mediante.

$$\begin{aligned} \Gamma &= \frac{2^{2C} - 1}{2^{2R} - 1} \\ &= \frac{\text{SNR}}{2^{2R} - 1} \end{aligned} \tag{6.193}$$

De modo equivalente, es posible escribir

$$R = \frac{1}{2}\log_2\left(1 + \frac{\text{SNR}}{\Gamma}\right) \quad \text{bits/transmisión} \tag{6.194}$$

Para una PAM o QAM codificada que opere a $P_e = 10^{-6}$, por ejemplo, la brecha Γ es constante a 8.8 dB. Mediante el uso de códigos (por ejemplo los códigos de enramado que se explican en el capítulo 10) la brecha Γ puede reducirse a un valor tan bajo como 1 dB.

Sea P la potencia de la señal transmitida y σ^2 la varianza del ruido del canal medida sobre el ancho de banda B. La relación señal a ruido es entonces

$$\text{SNR} = \frac{P}{\sigma^2}$$

donde

$$\sigma^2 = N_0 B$$

Por tanto, podemos definir finalmente la velocidad de datos alcanzable como

$$R = \frac{1}{2}\log_2\left(1 + \frac{P}{\Gamma\sigma^2}\right) \quad \text{bits/transmisión} \tag{6.195}$$

Con esta fórmula a nuestra disposición, estamos listos para describir la modulación multicanal en términos cuantitativos.

■ DIVISIÓN DEL CANAL EN TIEMPO CONTINUO

Considere un canal lineal de banda ancha (por ejemplo un par trenzado) con una respuesta en frecuencia arbitraria $H(f)$. Consideremos que la respuesta en magnitud al cuadrado

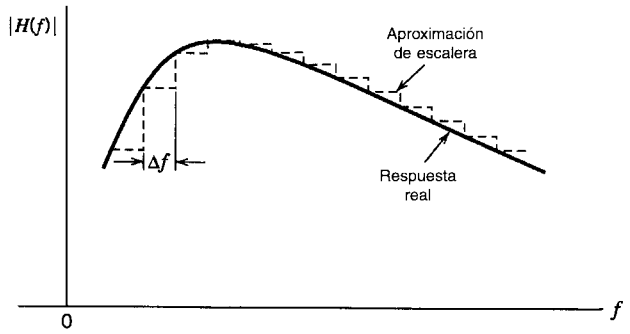


FIGURA 6.54 Aproximación de escalera de una respuesta en magnitud arbitraria $|H(f)|$; sólo se muestra la porción de frecuencia positiva de la respuesta.

$|H(f)|$ se aproxima por medio de una función de escalera, como se ilustra en la figura 6.54 con Δf denotando el ancho de cada escalón. En el límite, cuando el incremento de frecuencia Δf tiende a cero, la aproximación de escalera del canal se acerca a la $H(f)$ real. A lo largo de cada escalón de la aproximación es posible suponer que el canal opera como uno AWGN sin interferencia de intersímbolo. Por tanto, el problema de transmitir una sola señal de banda ancha se transforma en la transmisión de un conjunto de señales ortogonales de banda angosta. Cada una de éstas, con su propia portadora, se genera utilizando una técnica de modulación espectralmente eficiente, como la QAM M-aria, siendo el ruido blanco gaussiano aditivo en esencia la única fuente primaria de deterioro de transmisión, lo cual, a su vez, equivale a que la transmisión de datos por cada subcanal de ancho de banda Δf puede optimizarse apelando al teorema de la capacidad de información de Shannon, donde la optimización de cada subcanal la llevan a cabo de manera independiente de todos los demás. De ese modo, en términos prácticos del procesamiento de señales, *la necesidad de la ecualización complicada de un canal de banda ancha se reemplaza por la necesidad de multiplexar y demultiplexar la transmisión del flujo de datos entrante por un gran número de subcanales de banda angosta que son contiguos y disjuntos*. Aunque la complejidad resultante del sistema de multiportadora es en realidad elevada para un gran número de subcanales, es posible conseguir la puesta en práctica completa del sistema a un costo adecuado por medio del uso de la tecnología VLSI.

La figura 6.55 muestra un diagrama de bloques del sistema de transmisión de datos multicanal en su forma más básica. El sistema se configura aquí al utilizar la modulación por amplitud en cuadratura cuya elección se justifica por virtud de su eficiencia espectral. El flujo de datos binarios entrante se aplica primero a un demultiplexor (éste no se muestra en la figura), produciéndose de esa manera un conjunto de N subflujos. Cada uno de éstos representa una secuencia de *subsímbolos de dos elementos*, los cuales, en el intervalo del símbolo de $0 \le t \le T$, se denotan por medio de

$$(a_n, b_n), \quad n = 1, 2, \ldots, N$$

donde a_n y b_n son los valores de los elementos a lo largo de las dos coordenadas del subcanal n.

De manera correspondiente, las funciones básicas pasobanda de los moduladores en amplitud por cuadratura se definen mediante los pares de funciones

$$\{\phi(t)\cos(2\pi f_n t), \phi(t)\,\mathrm{sen}(2\pi f_n t)\}, \quad n = 1, 2, \ldots, N \tag{6.196}$$

donde la frecuencia f_n de la portadora del modulador n-ésimo es un múltiplo entero de la velocidad del símbolo $1/T$, de acuerdo con

$$f_n = \frac{n}{T}, \quad n = 1, 2, \ldots, N$$

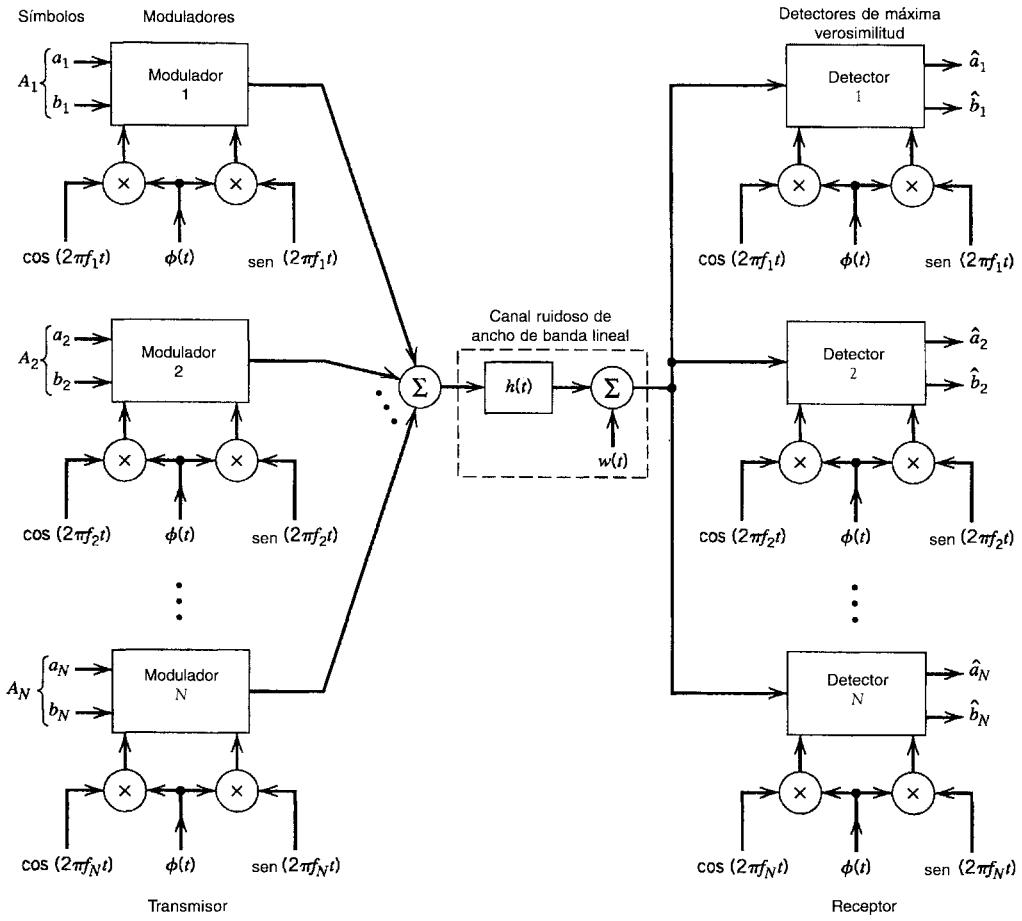


FIGURA 6.55 Diagrama de bloques de un sistema de transmisión de datos multicanal.

y la función $\phi(t)$ es la función senc:

$$\phi(t) = \sqrt{\frac{2}{T}}\text{senc}\left(\frac{t}{T}\right), \quad -\infty < t < \infty \tag{6.197}$$

Las funciones básicas pasobanda definidas aquí tienen las siguientes propiedades deseables (vea el problema 6.41 para sus demostraciones):

Propiedad 1

Para cada n, *las dos funciones senc moduladas en cuadratura forman un par ortogonal que se indica como*

$$\int_{-\infty}^{\infty} (\phi(t)\cos(2\pi f_n t))(\phi(t)\,\text{sen}(2\pi f_n t))\,dt = 0 \quad \text{para toda } n \tag{6.198}$$

Esta relación ortogonal proporciona la base para formular la constelación de señales correspondientes a cada uno de los N moduladores en la forma de una retícula cuadrada.

Propiedad 2

Admitiendo que

$$\exp(j2\pi f_n t) = \cos(2\pi f_n t) + j\,\mathrm{sen}(2\pi f_n t)$$

podemos redefinir completamente las funciones básicas pasobanda en la forma compleja

$$\left\{\frac{1}{\sqrt{2}}\phi(t)\exp(j2\pi f_n t)\right\}, \qquad n = 1, 2, \ldots, N \tag{6.199}$$

donde el factor $1/\sqrt{2}$ *se ha introducido para asegurar que la función escalada* $\phi(t)/\sqrt{2}$ *tenga energía unitaria. Por tanto, estas funciones básicas pasobanda constituyen un conjunto ortonormal, lo cual se indica mediante*

$$\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2}}\phi(t)\exp(j2\pi f_n t)\right)\left(\frac{1}{\sqrt{2}}\phi(t)\exp(j2\pi f_k t)\right)^* dt = \begin{cases}1, & k = n\\ 0, & k \neq n\end{cases} \tag{6.200}$$

donde el asterisco denota conjugación compleja.

La ecuación (6.200) proporciona la base matemática para asegurar que los N pares modulador–demodulador operen independientemente entre sí.

Propiedad 3

El conjunto de funciones de salida del canal $\{h(t) \bigstar \phi(t)\}$ *permanece ortogonal en un canal lineal con respuesta al impulso arbitraria* $h(t)$, *donde* $\bigstar$ *denota convolución.*

El canal se divide de este modo en un conjunto de subcanales independientes que operan en tiempo continuo.

La figura 6.55 incluye también la estructura del receptor. Ésta consiste en un banco de N detectores coherentes, con la salida aplicándose en forma simultánea en las entradas del detector. Cada detector se alimenta con un par generado localmente de funciones senc moduladas en cuadratura que operan en sincronía con el par de funciones básicas pasobanda aplicadas al correspondiente modulador en el transmisor.

Cada subcanal puede tener alguna interferencia de intersímbolo residual. Sin embargo, cuando el número de subcanales N tiende a infinito, dicha interferencia desaparece. De tal manera, para una N suficientemente grande, el banco de detectores coherentes en la figura 6.55 opera como *detectores de máxima verosimilitud*, independientemente uno de otro y sobre un esquema de subsímbolo a subsímbolo.

Para definir las salidas del detector en respuesta a los subsímbolos de entrada, encontramos conveniente utilizar la notación compleja. Sea A_n el subsímbolo aplicado al modulador n-ésimo durante el intervalo de símbolo $0 \leq t \leq T$:

$$A_n = a_n + jb_n, \quad n = 1, 2, \ldots, N \tag{6.201}$$

La salida del detector correspondiente es

$$Y_n = H_n A_n + W_n, \quad n = 1, 2, \ldots, N \tag{6.202}$$

donde H_n es la respuesta en frecuencia de valor complejo del canal evaluada a la frecuencia de portadora del subcanal $f = f_n$:

$$H_n = H(f_n), \quad n = 1, 2, \ldots, N \tag{6.203}$$

La W_n es la variable aleatoria de valor compleja debida al ruido del canal $w(t)$; las partes real e imaginaria de W_n tienen media cero y varianza $N_0/2$. Al conocer la respuesta en frecuencia

medida $H(f)$, podemos usar entonces la ecuación (6.202) para calcular una estimación de máxima verosimilitud del subsímbolo transmitido A_n. Las estimaciones $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_N$ obtenidas de ese modo se multiplexan finalmente para producir la estimación correspondiente de los datos binarios originales transmitidos durante el intervalo $0 \leq t \leq T$.

En resumen, para una N suficientemente grande, es posible poner en práctica el receptor como un detector de máxima verosimilitud óptimo, que opere como N detectores de subsímbolo por subsímbolo. La razón por la que es posible construir un receptor de máxima verosimilitud de una manera tan simple radica en que las funciones básicas pasobanda constituyen un conjunto ortonormal, y su ortogonalidad se mantiene para toda respuesta al impulso $h(t)$ del canal.

■ RELACIÓN GEOMÉTRICA SEÑAL A RUIDO

En el sistema de transmisión multicanal de la figura 6.55, cada subcanal está caracterizado por una SNR propia. Sería muy deseable obtener una sola medida del desempeño de la figura 6.55.

Para simplificar la deducción de una medida de este tipo, suponemos que todos los subcanales en la figura 6.55 se representan por medio de constelaciones unidimensionales. En ese caso, la capacidad del canal del sistema completo en bits por transmisión está dada

$$\begin{aligned} R &= \frac{1}{N}\sum_{n=1}^{N} R_n \\ &= \frac{1}{2N}\sum_{n=1}^{N} \log_2\left(1+\frac{P_n}{\Gamma\sigma_n^2}\right) \\ &= \frac{1}{2N}\log_2 \prod_{n=1}^{N}\left(1+\frac{P_n}{\Gamma\sigma_n^2}\right) \\ &= \frac{1}{2}\log_2\left[\prod_{n=1}^{N}\left(1+\frac{P_n}{\Gamma\sigma_n^2}\right)\right]^{1/N} \end{aligned} \tag{6.204}$$

Sea $(\text{SNR})_{\text{total}}$ la relación señal a ruido total del sistema completo. Es posible expresar entonces R bits por transmisión como:

$$R = \frac{1}{2}\log_2\left(1+\frac{(\text{SNR})_{\text{total}}}{\Gamma}\right) \tag{6.205}$$

Comparando las ecuaciones (6.205) con la (6.204), podemos escribir

$$(\text{SNR})_{\text{total}} = \Gamma\left(\prod_{n=1}^{N}\left(1+\frac{P_n}{\Gamma\sigma_n^2}\right)^{1/N} - 1\right) \tag{6.206}$$

Suponiendo que $P_n/\Gamma\sigma_n^2$ es lo suficientemente alta para ignorar los dos términos unidad en la ecuación (6.206), es posible aproximar la SNR total como

$$(\text{SNR}) = \prod_{n=1}^{N}\left(\frac{P_n}{\sigma_n^2}\right)^{1/N} \tag{6.207}$$

Podemos caracterizar de este modo al sistema completo mediante una SNR que es la media geométrica de las SNR de los subcanales individuales.

La SNR geométrica de la ecuación (6.207) puede mejorarse de manera considerable distribuyendo la potencia de transmisión disponible entre los N subcanales de acuerdo con una base no uniforme. Este objetivo se logra mediante el empleo de la carga, como se expone a continuación.

■ CARGA DEL SISTEMA DE TRANSMISIÓN MULTICANAL

La ecuación (6.204) para la velocidad de bits del sistema multicanal completo ignora el efecto del canal en el desempeño del sistema. Para tomar en cuenta dicho efecto, se define

$$g_n = |H(f_n)|, \quad n = 1, 2, \ldots, N \tag{6.208}$$

Suponiendo en ese caso que el número de subcanales N es bastante grande, es posible considerar que g_n es constante sobre el ancho de banda total Δf asignado al subcanal n para toda n. En tal caso, modificamos la segunda línea de la ecuación (6.204) para la SNR total del sistema como

$$R = \frac{1}{2N} \sum_{n=1}^{N} \log_2 \left(1 + \frac{g_n^2 P_n}{\Gamma \sigma_n^2} \right) \tag{6.209}$$

La g_n^2 y Γ suelen ser fijas. La varianza del ruido σ_n^2 *es* $\Delta f N_0$ para toda n, donde Δf es el ancho de banda de cada subcanal y $N_0/2$ es la densidad espectral de potencia del ruido. Por tanto, podemos optimizar la velocidad de bits completa R mediante una asignación adecuada de la potencia de transmisión total entre los diversos canales. Sin embargo, para que esta optimización tenga un valor práctico, debemos mantener la potencia de transmisión total en cierto valor constante P, digamos, como se indica mediante

$$\sum_{n=1}^{N} P_n = P = \text{constante} \tag{6.210}$$

La optimización con la que tenemos que tratar en ese caso constituye un *problema de optimización restringido,* que se establece como sigue:

> Maximizar la velocidad de bits R para el sistema de transmisión multicanal completo al compartir de manera óptima la potencia de transmisión total P entre los N subcanales de acuerdo con la restricción de que P se mantiene constante.

Para resolver este problema de optimización utilizamos primero el *método de los multiplicadores de Lagrange*[19] para establecer una función objetivo que incorpora la restricción de la ecuación (6.210), como se indica a partir de

$$\begin{aligned} J &= \frac{1}{2N} \sum_{n=1}^{N} \log_2 \left(1 + \frac{g_n^2 P_n}{\Gamma \sigma_n^2} \right) + \lambda \left(P - \sum_{n=1}^{N} P_n \right) \\ &= \frac{1}{2N} \log_2 e \sum_{n=1}^{N} \log_e \left(1 + \frac{g_n^2 P_n}{\Gamma \sigma_n^2} \right) + \lambda \left(P - \sum_{n=1}^{N} P_n \right) \end{aligned} \tag{6.211}$$

donde λ es el *multiplicador de Lagrange*. Por tanto, al diferenciar J con respecto a P_n, igualando después a cero el resultado y arreglando por último los términos, obtenemos

$$\frac{\frac{1}{2N} \log_2 e}{P_n + \frac{\Gamma \sigma_n^2}{g_n^2}} = \lambda \tag{6.212}$$

Este resultado indica que la solución a nuestro problema de optimización restringido radica en tener

$$P_n + \frac{\Gamma \sigma_n^2}{g_n^2} = K \quad \text{para } n = 1, 2, \ldots, N \tag{6.213}$$

donde K es una constante preestablecida bajo el control del diseñador. Es decir, la suma de la potencia de transmisión y de la varianza del ruido (potencia) escalada por el cociente Γ/g_n^2 debe mantenerse constante en cada subcanal. El proceso de asignar la potencia transmitida P a los subcanales individuales de modo que se maximice la velocidad de bits del sistema de transmisión multicanal completo, recibe el nombre de *carga*.

■ INTERPRETACIÓN DE LA ANALOGÍA DEL LLENADO DE AGUA PARA EL PROBLEMA DE OPTIMIZACIÓN

Al resolver el problema de optimización restringido que se acaba de describir, deben satisfacerse dos condiciones, a saber, las ecuaciones (6.210) y (6.213). La solución óptima definida de esa manera tiene una interesante interpretación, como se ilustra en la figura 6.56 para $N = 6$, suponiendo que la brecha Γ es constante sobre todos los subcanales. Para simplificar la ilustración en la figura 6.56 hemos dejado $\sigma_n^2 = N_0 \Delta f = 1$, es decir, la potencia promedio del ruido es la unidad en todos los subcanales N. Con referencia a esta figura, podemos señalar ahora las siguientes observaciones:

- La suma de la potencia P_n asignada al canal n y la potencia del ruido escalado Γ/g_n^2 satisface la restricción de la ecuación (6.213) en cuatro de los subcanales para un valor preestablecido de la constante K.
- La suma de las asignaciones de potencia a estos cuatro subcanales consume toda la potencia de transmisión disponible, mantenida a un valor constante P.
- Los dos subcanales restantes se han dejado de considerar debido a que requerirían potencia negativa para satisfacer la ecuación (6.213) para el valor preestablecido de la constante K. Esta condición es evidentemente inaceptable.

La interpretación que se ilustra en la figura 6.56 nos lleva a referirnos a la solución óptima de la ecuación (6.213), sujetos a la restricción de la ecuación (6.210), como la *solución del llenado de agua*. Esta terminología se desprende de la analogía de nuestro problema de optimización con una cantidad fija de agua (que representa a la potencia de transmisión) que se está vertiendo dentro de un recipiente con varias regiones conectadas, cada una con profundidades diferentes (lo que representa a la potencia del ruido). El agua se distribuye por sí sola de tal manera que se logra un nivel de agua constante en el recipiente completo. En el capítulo 9 abundamos sobre la interpretación del llenado de agua con respecto a la capacidad de información.

Volviendo a la tarea de cómo asignar la potencia de transmisión fija P entre los diversos subcanales de un sistema de transmisión multicanal de modo que se optimice la velocidad de bits del

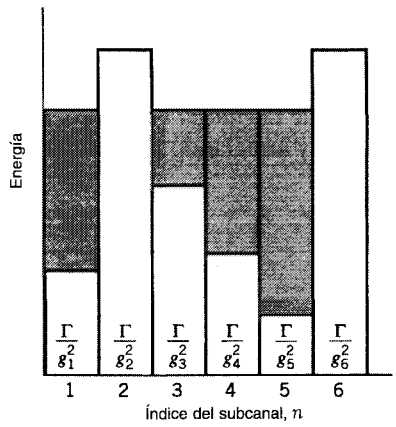


FIGURA 6.56 Interpretación de la analogía del llenado de agua para el problema de carga.

sistema completo, es posible proceder como sigue. Considérese que la potencia de transmisión total se fija en un valor constante P, como en la ecuación (6.210). Sea K el valor constante preestablecido para la suma $P_n + \Gamma\sigma_n^2 / g_n^2$ para toda n, como en la ecuación (6.213). Podemos utilizar entonces este par de ecuaciones para integrar el siguiente sistema de ecuaciones simultáneas:

$$\begin{aligned} P_1 + P_2 + \ldots P_N &= P \\ P_1 - K &= -\Gamma\sigma^2 / g_1^2 \\ P_2 - K &= -\Gamma\sigma^2 / g_2^2 \\ &\vdots \\ P_N - K &= -\Gamma\sigma^2 / g_N^2 \end{aligned} \tag{6.214}$$

donde tenemos un total de $(N + 1)$ incógnitas y $(N + 1)$ ecuaciones por resolver con respecto a ellas. Es posible reescribir este conjunto de ecuaciones simultáneas en forma de matriz como:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ 1 & 0 & \cdots & 0 & -1 \\ 1 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \\ K \end{bmatrix} = \begin{bmatrix} P \\ -\Gamma\sigma^2 / g_1^2 \\ -\Gamma\sigma^2 / g_2^2 \\ \vdots \\ -\Gamma\sigma^2 / g_N^2 \end{bmatrix} \tag{6.215}$$

Al premultiplicar ambos miembros de la ecuación (6.215) por la inversa de la matriz de $(N + 1)$ por $(N + 1)$ del primer miembro de la ecuación, obtenemos soluciones para las incógnitas P_1, $P_2, \ldots, P_N$ y K. Debemos encontrar siempre que K es positiva, aunque es posible para algunas otras P que sea negativa. Las P negativas se descartan porque la potencia no puede ser negativa.

Ejemplo 6.7

Considere un canal lineal cuya respuesta en magnitud al cuadrado $|H(f)|^2$ tiene la forma lineal por secciones que se muestra en la figura 6.57. Para simplificar el ejemplo establecemos la brecha $\Gamma = 1$ y la varianza $\sigma^2 = 1$. En la situación así descrita, la aplicación de la ecuación (6.214) produce

$$\begin{aligned} P_1 + P_2 &= P \\ P_1 - K &= -1 \\ P_2 - K &= -1/l \end{aligned}$$

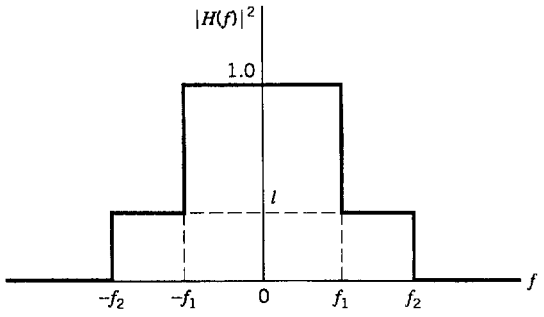


Figura 6.57 Respuesta en magnitud al cuadrado para el ejemplo 6.7.

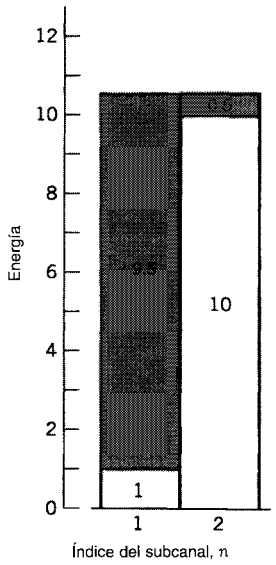


FIGURA 6.58 Perfil del llenado de agua para el ejemplo 6.7.

donde la potencia de transmisión total P se normaliza con respecto a la varianza del ruido. Al resolver estas tres ecuaciones simultáneas para P_1, P_2 y K, obtenemos

$$P_1 = \frac{1}{2}\left(P - 1 + \frac{1}{l}\right)$$
$$P_2 = \frac{1}{2}\left(P + 1 - \frac{1}{l}\right)$$
$$K = \frac{1}{2}\left(P + 1 + \frac{1}{l}\right)$$

Puesto que $0 < l < 1$, se concluye que $P_1 > 0$, pero es posible que P_2 sea negativa. Esta última condición tal vez surja si

$$l < \frac{1}{P+1}$$

Aunque en ese caso P_1 supera el valor preestablecido de la potencia transmitida P. Por tanto, se concluye que en este ejemplo la única solución aceptable es tener $1/(P + 1) < l < 1$. Suponga entonces que tenemos $P = 10$ y $l = 0.1$, para lo cual la solución corresponde a

$$K = 10.5$$
$$P_1 = 9.5$$
$$P_2 = 0.5$$

La imagen correspondiente de la analogía del llenado del agua se presenta en la figura 6.58. ◀

6.13 *Multitonos discretos*

El material que se presenta en la sección 6.12 proporciona una introducción sumamente informativa para la noción de modulación multicanal. En particular, la división del canal en tiempo continuo inducida por las funciones básicas pasobanda de la ecuación (6.196) o equivalentemente (6.199) exhibe una propiedad muy deseable: la ortogonalidad de las funciones básicas (y, por

ello, la división del canal) se preserva a pesar de su convolución con la respuesta al impulso del canal. Sin embargo, el sistema tiene dos desventajas:

1. Las funciones básicas pasobanda utilizan una función senc que es distinta de cero para un intervalo de tiempo finito, siempre que las consideraciones prácticas favorezcan un intervalo de observación finito.
2. Para un número finito de subcanales, N, el sistema es subóptimo; la optimización del sistema se asegura sólo cuando N tiende a infinito.

Podemos superar estas desventajas utilizando *multitono discreto* (DMT), cuya idea básica es transformar un canal de banda ancha en un conjunto de N subcanales que operan en paralelo. Lo que distingue al DMT es el hecho de que la transformación se realiza en tiempo discreto, así como en frecuencia discreta. En consecuencia, el comportamiento de entrada-salida del transmisor del sistema de comunicación completo admite una representación de matriz lineal, que por sí sola tiende a la puesta en práctica utilizando la transformada discreta.

Para explorar este nuevo método, admitimos primero que en una situación realista el canal tiene su respuesta al impulso distinta de cero, $h(t)$, esencialmente confinada a un intervalo finito $[0, T_b]$. Así, dejamos que la secuencia $h_0, h_1, \ldots, hv$ denote la respuesta al impulso equivalente de banda base del canal muestreado a una tasa $1/T_s$, con

$$T_b = (1+v)T_s \tag{6.216}$$

La velocidad de muestreo $1/T_s$ se elige mayor que el doble de la componente de frecuencia más alta de interés, de acuerdo con el teorema del muestreo. Para continuar con la descripción en tiempo discreto del sistema, sea $s[n] = s(nT_s)$ una muestra del símbolo transmitido $s(t)$, $w[n] = w(nT_s)$ una muestra del ruido del canal $w(t)$, y $x[n] = x(nT_s)$ la muestra correspondiente de la salida del canal (señal recibida). El canal realiza convolución lineal sobre la secuencia de símbolos entrante $\{s[n]\}$ de longitud N, produciendo una secuencia de salida del canal $\{x[n]\}$ de longitud $N + v$. La extensión de la secuencia de salida del canal en v muestras en comparación con la secuencia de entrada se debe a la interferencia de intersímbolo que produce el canal.

Para superar el efecto de la interferencia de intersímbolo, creamos un *intervalo de seguridad* extendido cíclicamente, en el cual a cada secuencia de símbolos precede una extensión periódica de la secuencia misma. De manera específica, las últimas v muestras de la secuencia de símbolos se repiten al principio de la secuencia que se transmite, como indica

$$s[-k] = s[N-k] \quad \text{para } k = 1, 2, \ldots, v \tag{6.217}$$

Esta condición se denomina un *prefijo cíclico*. El *factor de ancho de banda de exceso* debido a la inclusión de un prefijo cíclico es, por tanto, v/N, donde N es el número de muestras transmitidas después del intervalo de seguridad.

Con el prefijo cíclico aplicado, la descripción del canal toma la forma

$$\begin{bmatrix} x[N-1] \\ x[N-2] \\ \vdots \\ x[N-v-1] \\ x[N-v-2] \\ \vdots \\ x[0] \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_{v-1} & h_v & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{v-2} & h_{v-1} & h_v & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_v \\ h_v & 0 & 0 & \cdots & 0 & 0 & h_0 & \cdots & h_{v-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ h_1 & h_2 & h_3 & \cdots & h_v & 0 & 0 & \cdots & h_0 \end{bmatrix} \begin{bmatrix} s(N-1) \\ s(N-2) \\ \vdots \\ s[N-v-1] \\ s[N-v-2] \\ \vdots \\ s[0] \end{bmatrix} + \begin{bmatrix} w[N-1] \\ w[N-2] \\ \vdots \\ w[N-v-1] \\ w[N-v-2] \\ \vdots \\ w[0] \end{bmatrix} \tag{6.218}$$

De manera equivalente, es posible describir la representación en tiempo discreto del canal en la forma matriz compacta

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{6.219}$$

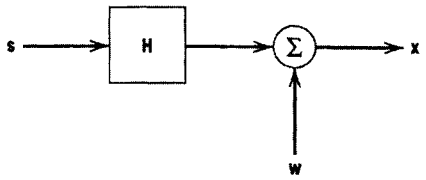


FIGURA 6.59 Representación en tiempo discreto del sistema de transmisión de datos multicanal.

donde el vector del símbolo transmitido **s**, el vector del ruido del canal **w** y el vector de la señal recibida **x** son todos vectores de N por 1, que se definen respectivamente por medio de

$$\mathbf{s} = [s[N-1], s[N-2], \ldots, s[0]]^T \tag{6.220}$$

$$\mathbf{w} = [w[N-1], w[N-2], \ldots, w[0]]^T \tag{6.221}$$

y

$$\mathbf{x} = [x[N-1], x[N-2], \ldots, x[0]]^T \tag{6.222}$$

Por tanto, podemos describir la representación en tiempo discreto del canal como en la figura 6.59. La matriz **H** de N por N del canal se define mediante

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_{v-1} & h_v & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{v-2} & h_{v-1} & h_v & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_v \\ h_v & 0 & 0 & \cdots & 0 & 0 & h_0 & \cdots & h_{v-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ h_1 & h_2 & h_3 & \cdots & h_v & 0 & 0 & \cdots & h_0 \end{bmatrix} \tag{6.223}$$

A partir de esta definición, vemos de inmediato que la matriz **H** tiene la siguiente composición estructural: cada renglón de la matriz se obtiene aplicando un corrimiento de una posición hacia la derecha del renglón anterior, con la salvedad agregada de que el elemento más a la derecha del renglón anterior se cae en el proceso de corrimiento y "circula" hasta el elemento más hacia la izquierda del nuevo renglón. En consecuencia, **H** recibe el nombre de *matriz circulante*.

Antes de seguir, conviene revisar brevemente la transformada discreta de Fourier y su papel en la descomposición espectral de la matriz circulante **H**.

■ TRANSFORMADA DISCRETA DE FOURIER

Considere el vector **x** de N por 1 de la ecuación (6.222). La transformada discreta de Fourier (DFT) del vector **x** se define a partir del vector de N por 1

$$\mathbf{X} = [X[N-1], X[N-2], \ldots, X[0]]^T \tag{6.224}$$

donde

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N-1} x[n] \exp\left(-j\frac{2\pi}{N}kn\right), \qquad k = 0, 1, \ldots, N-1 \tag{6.225}$$

El término exponencial $\exp(-j2\pi kn/N)$ se denomina el *kernel* o núcleo de la DFT. En forma correspondiente, la *transformada discreta de Fourier inversa (IDFT)* del vector **X** de N por 1 está definida por

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] \exp\left(j\frac{2\pi}{N}kn\right), \qquad n = 0, 1, \ldots, N-1 \tag{6.226}$$

Aun cuando las ecuaciones (6.225) y (6.226) parecen similares, tienen diferentes interpretaciones. Dado el vector de la señal **x**, la ecuación (6.225) proporciona una representación espectral de la señal calculada en un conjunto de frecuencias discretas: $f_k = k/N$, las cuales están normalizadas con respecto a la velocidad de muestreo. Dado el vector transformado **X**, la ecuación (6.226) recupera al vector de señal original **x**. Por tanto, podemos considerar la expresión (6.225) como la *ecuación de análisis* y a la (6.226) como la *ecuación de síntesis*.

Una propiedad de importancia de una matriz circulante, ejemplificada por la matriz **H** del canal de la ecuación (6.223), es que permite la *descomposición espectral*, como indica

$$\mathbf{H} = \mathbf{Q}^{\dagger}\mathbf{\Lambda}\mathbf{Q} \tag{6.227}$$

donde el exponente daga (†) denota la *transposición hermitiana* (es decir, la combinación de conjugación compleja y la transposición de matrices ordinaria). Las descripciones de las matrices **Q** y **Λ** se representan a continuación en ese orden.

La matriz **Q** es una matriz cuadrada definida en términos del kernel de la DFT de N puntos en la forma siguiente:

$$\mathbf{Q} = \frac{1}{\sqrt{N}}\begin{bmatrix} \exp\left(-j\frac{2\pi}{N}(N-1)(N-1)\right) & \cdots & \exp\left(-j\frac{2\pi}{N}2(N-1)\right) & \exp\left(-j\frac{2\pi}{N}(N-1)\right) & 1 \\ \exp\left(-j\frac{2\pi}{N}(N-1)(N-2)\right) & \cdots & \exp\left(-j\frac{2\pi}{N}2(N-2)\right) & \exp\left(-j\frac{2\pi}{N}(N-2)\right) & 1 \\ \vdots & & \vdots & \vdots & \vdots \\ \exp\left(-j\frac{2\pi}{N}(N-1)\right) & \cdots & \exp\left(-j\frac{2\pi}{N}2\right) & \exp\left(-j\frac{2\pi}{N}\right) & 1 \\ 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \tag{6.228}$$

A partir de esta definición, vemos de inmediato que el elemento k-ésimo de la matriz de N por N, **Q**, empezando desde la *parte inferior derecha* en $k = 0$ y $l = 0$ y continuando hacia arriba en forma gradual, es

$$q_{kl} = \frac{1}{\sqrt{N}}\exp\left(-j\frac{2\pi}{N}kl\right), \quad (k,l) = 0,1,\ldots,N-1 \tag{6.229}$$

La matriz **Q** es una *matriz ortogonal* o *matriz unitaria* con respecto a que satisface la condición

$$\mathbf{Q}^{\dagger}\mathbf{Q} = \mathbf{I} \tag{6.230}$$

donde **I** es la matriz identidad. Es decir, la matriz inversa de **Q** es igual a la transpuesta hermitiana de **Q**.

La matriz **Λ** es una *matriz diagonal* que contiene los N valores de la transformada discreta de Fourier de la secuencia $h_0, h_1, \ldots, h_v$ que caracterizan al canal. Denotando estos valores de la transformada mediante $\lambda_{N-1}, \ldots, \lambda_1, \lambda_0$, podemos expresar **Λ** como

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{N-1} & 0 & \cdots & 0 \\ 0 & \lambda_{N-2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_0 \end{bmatrix} \tag{6.231}$$

(Las λ aquí no deben confundirse con los multiplicadores de Lagrange de la sección 6.12.)

La DFT se ha establecido por sí sola como una de las herramientas principales del procesamiento de señales digitales, en virtud de su eficiente cálculo utilizando el *algoritmo de la trans-*

formada rápida de Fourier (FFT).[20] De modo específico, el algoritmo FFT requiere del orden de $N \log_2 N$ operaciones más que las operaciones N^2 para el cálculo directo de las DFT. Para la puesta en práctica eficiente del algoritmo FFT, debemos elegir la longitud del bloque N como una potencia entera de dos. Los ahorros de cálculo que se obtienen al usar el algoritmo FFT se consiguen al explotar la estructura especial DFT que se define en la ecuación (6.225). Además, estos ahorros se vuelven más sustanciales cuando aumentamos la longitud N de los datos.

■ DESCRIPCIÓN DEL CANAL EN EL DOMINIO DE LA FRECUENCIA

Con la breve descripción de la DFT, estamos listos para concluir nuestra exposición de los multitonos discretos. En primer lugar definimos

$$\mathbf{s} = \mathbf{Q}^{\dagger}\mathbf{S} \tag{6.232}$$

donde $\mathbf{S}$ es la representación vectorial en el dominio de la frecuencia de la entrada del transmisor. Cada elemento del vector $\mathbf{S}$ de N por uno puede considerarse como un punto de valor complejo en una constelación de señales QAM bidimensional. Dado el vector $\mathbf{x}$ de salida del canal, definimos su representación correspondiente en el dominio de la frecuencia como

$$\mathbf{X} = \mathbf{Q}\mathbf{x} \tag{6.233}$$

Mediante el uso de las ecuaciones (6.227), (6.232) y (6.233) es posible reescribir la ecuación (6.219) en la forma equivalente

$$\mathbf{X} = \mathbf{Q}(\mathbf{Q}^{\dagger}\mathbf{\Lambda}\mathbf{Q}\mathbf{Q}^{\dagger}\mathbf{S} + \mathbf{W}) \tag{6.234}$$

Por tanto, utilizando la relación de la ecuación (6.230), obtenemos simplemente

$$\mathbf{X} = \mathbf{\Lambda}\mathbf{S} + \mathbf{W} \tag{6.235}$$

donde

$$\mathbf{W} = \mathbf{Q}\mathbf{w} \tag{6.236}$$

En forma expandida, la ecuación (6.235) se lee como

$$X_k = \lambda_k S_k + W_k, \quad k = 0, 1, \ldots, N-1 \tag{6.237}$$

donde el conjunto de valores en el dominio de la frecuencia $\{\lambda_k\}_{k=0}^{N-1}$ se conoce para un canal preestablecido.

Para un canal con ruido blanco aditivo, la ecuación (6.237) implica que el receptor está compuesto de un conjunto de procesadores independientes que operan en paralelo. Conociendo todas las λ_k, podemos utilizar el bloque de valores en el dominio de la frecuencia $\{X_k\}_{k=0}^{N-1}$ para calcular estimaciones del correspondiente bloque transmitido de valores en el dominio de la frecuencia $\{S_k\}_{k=0}^{N-1}$.

■ SISTEMA DMT BASADO EN LA DFT

Las ecuaciones (6.235), (6.225), (6.226) y (6.237) dan las bases matemáticas para la puesta en práctica de la DMT utilizando la DFT. La figura 6.60 ilustra el diagrama de bloques del sistema que se obtiene de estas ecuaciones y sus implicaciones prácticas.

El transmisor está compuesto por los siguientes bloques funcionales:

- *Demultiplexor que convierte el flujo de datos en serie entrante en forma paralelo.*
- *Codificador de constelación*, que hace corresponder los datos en paralelo en $N/2$ subcanales multibit con cada subcanal que está representando la constelación de señales QAM. La asignación de bits entre los subcanales también se efectúa aquí de acuerdo con el algoritmo de carga.

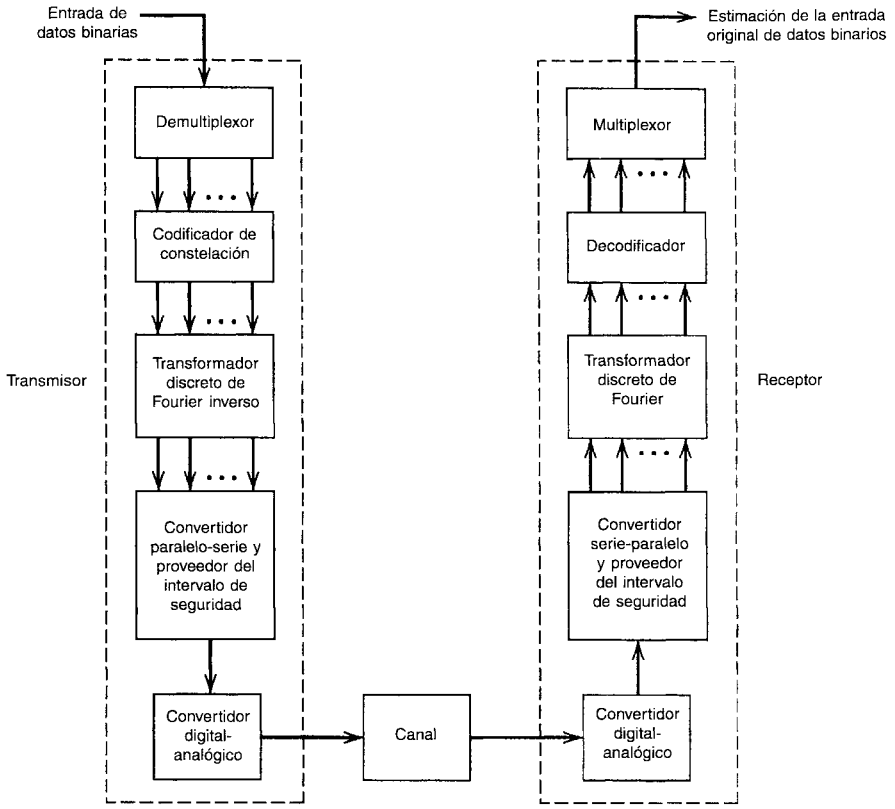


FIGURA 6.60 Diagrama de bloques de un sistema de transmisión de datos de multitonos discretos (DMT).

- *Transformador discreto de Fourier inverso (IDFT)*, que transforma los datos en paralelo en el dominio de la frecuencia a la salida del codificador de constelación en datos en paralelo en el dominio del tiempo. Para la puesta en práctica eficiente de la IDFT utilizando el algoritmo de la transformada rápida de Fourier (FFT), necesitamos elegir $N = 2^k$, donde k es un entero positivo.
- *Convertidor paralelo-serie*, que convierte los datos en paralelo en el dominio del tiempo en la forma en serie. Se insertan intervalos de seguridad rellenados con prefijos cíclicos en los datos en serie con base en un esquema periódico antes de la conversión en la forma analógica.
- *Convertidor digital-analógico (DAC)*, que convierte los datos digitales en forma analógica lista para la transmisión por el canal.

Por lo común, el convertidor digital-analógico incluye un filtro de transmisión. Por tanto, la función de tiempo $h(t)$ debe redefinirse como la respuesta al impulso combinada de la conexión en cascada del filtro de transmisión y el canal.

El receptor realiza las operaciones inversas del transmisor, como se describe enseguida:

- *Convertidor analógico-digital (ADC)*; convierte la salida analógica del canal en digital.
- *Convertidor serie-paralelo*, que transforma el flujo de bits que se produce en forma paralela. Antes de que se realice esta conversión se eliminan los intervalos de seguridad (prefijos cíclicos).
- *Transformador discreto de Fourier (DFT)*, el cual transforma los datos paralelos en el dominio del tiempo en datos paralelos en el dominio de la frecuencia; como en el caso de las IDFT, el algoritmo FFT se usa para poner en práctica la transformada discreta de Fourier.

- *Decodificador,* el cual utiliza la DFT para calcular estimaciones de los datos de subcanal multibit originales que se alimentan al transmisor.
- *Multiplexor*, que combina las estimaciones así calculadas para reproducir una reconstrucción de flujo de datos transmitidos en serie.

■ APLICACIONES DE LOS DMT

Una importante aplicación de los multitonos discretos corresponde a la transmisión de datos por canales de dos vías. En realidad, la DMT se ha estandarizado para utilizarse en *líneas digitales asimétricas de suscriptor* (ADSL) empleando pares trenzados. Las ADSL se describieron en el capítulo 4. Por ejemplo, los DMT se ofrecen para la transmisión de datos directa (es decir, desde un proveedor de servicios de Internet hasta un suscriptor) a la velocidad DS1 de 1.544 Mb/s y la transmisión simultánea de datos en dirección inversa (es decir, desde el suscriptor hasta el proveedor de servicios de Internet) a 160 kb/s. Este tipo de capacidad de transmisión de datos resulta bastante adecuada para manejar aplicaciones intensivas en datos, como el video sobre demanda.

Los DMT constituyen también una tecnología fundamental en la puesta en práctica de las *líneas de suscriptor digitales de muy alta velocidad*[21] (VDSL), que difieren de todas las demás técnicas de transmisión DSL debido a su capacidad para entregar velocidades de datos extremadamente elevadas. Por ejemplo, la VDSL puede proporcionar velocidades de datos de 13 a 26 Mb/s en dirección directa y de 2 a 3 Mb/s en la inversa por pares trenzados que emanan de una unidad de red óptica y que conectan al suscriptor a lo largo de distancias menores que aproximadamente 1 km. Estas altas velocidades de datos nos permiten la entrega de TV digital, navegación en la web y transferencia de archivos super rápidas, así como oficinas virtuales en casa.

El uso de los multitonos digitales para las líneas digitales asimétricas de suscriptor o de muy alta velocidad proporciona numerosas ventajas:

- *La capacidad de maximizar la tasa de bits transmitidos*, lo cual se proporciona adaptando la distribución de las señales que contienen información a través del canal, de acuerdo con la atenuación de este mismo y de las condiciones de ruido.
- *Adaptabilidad a las condiciones cambiantes de la línea*, la cual se ejecuta porque el canal se divide en varios subcanales.
- *Sensibilidad reducida ante el ruido de impulso,* lo cual se logra dispersando su energía por los diversos subcanales del receptor. Como su nombre lo indica, el *ruido de impulso* se caracteriza por largos intervalos de silencio seguidos por pulsos angostos de amplitud que varía al azar. En los ambientes de ADSL o VDSL, el ruido de impulso surge como resultado de transitorios de conmutación acoplados a pares de alambre en la oficina central y a diversos dispositivos eléctricos en las instalaciones del usuario.

■ COMPARACIÓN ENTRE LÍNEAS DE SUSCRIPCIÓN DIGITAL Y MÓDEMS DE BANDA DE VOZ

En la sección 6.11 explicamos los módems de la banda de voz que ya se encuentran cerca de operar en sus límites teóricos de 33.6 kb/s en dirección inversa y de 56 kb/s en la directa. En esta sección hemos explicado la aplicación de los DMT en las VDSL que pueden operar a velocidades de datos de casi 2 a 3 Mb/s en dirección inversa y de 13 a 26 Mb/s en la directa. Estos dos conjuntos bastante diferentes de velocidades de datos en dirección inversa/directa conducen a la siguiente interrogante: ¿cómo es posible que las VDSL operen a velocidades casi tres órdenes de magnitud más rápidas que los módems de la banda de voz por los mismos pares trenzados (es decir, líneas telefónicas)? La razón de esta enorme diferencia en las velocidades de datos de operación entre los módems de banda de voz y las VDSL no se encuentra en los pares trenzados, más bien radica en los *interruptores digitales* incorporados en la red telefónica pública conmutable que evita el transporte de datos de banda ancha a los suscriptores (usuarios) por medio de los módems de banda de voz. Planteándolo en términos simples, los interruptores digitales tratan a los datos digitales de la misma manera que las señales de voz para las cuales se diseñaron fundamentalmente.

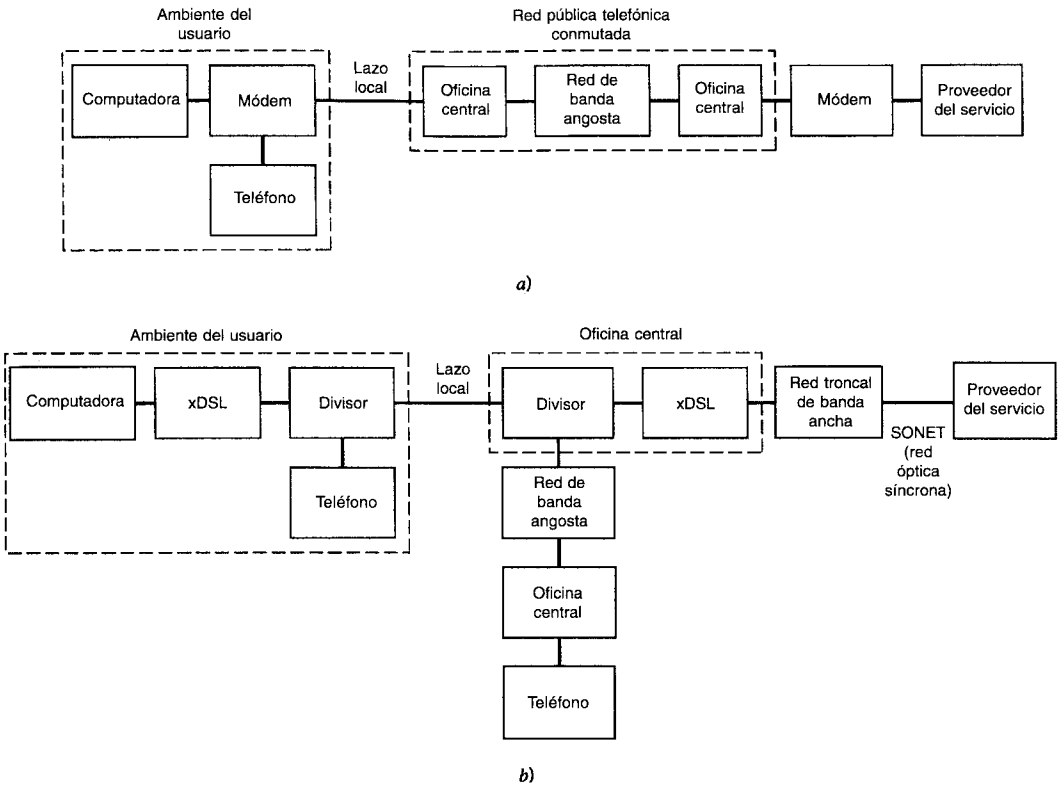


FIGURA 6.61 *a*) Ambiente del módem de banda de voz. *b*) Ambiente xDSL (línea de suscriptor digital), donde x significa "asimétrico" o "muy alta velocidad".

La figura 6.61 destaca el ambiente operativo de los modos de la banda de voz y las xDSL, donde x representa a la A en ADSL y a la V en VDSL. En el modelo de la figura 6.61*a* relativo al módem de la banda de voz, tenemos una trayectoria de transmisión relativamente larga entre un proveedor de servicios de Internet (ISP)y un suscriptor, pero es más importante el hecho de que la trayectoria de transmisión recorre una red pública de conmutación telefónica (PSTN), que limita el ancho de banda de canal disponible a casi 3.5 kHz. En contraste, en el modelo de la figura 6.61*b* correspondiente a las xDSL, la trayectoria de transmisión aloja el transporte de los datos de banda ancha entre el ISP y el suscriptor por medio de una red digital de servicios integrados de banda ancha y de un lazo local relativamente corto que se basa en un par trenzado. El sistema permite la coexistencia del servicio telefónico convencional y las señales xDSL en el mismo lazo local, lo cual se hace posible a través del uso de un par de divisores, como se indica en la figura 6.61*b*; los divisores, consistentes en filtros bidireccionales pasobajas y pasoaltas, se explican en la sección 4.8.

■ MULTIPLEXADO ORTOGONAL POR DIVISIÓN DE FRECUENCIA[22]

Los multitonos discretos constituyen una forma discreta particular de la modulación multicanal. Otra forma muy relacionada de este método de modulación es el *multiplexado ortogonal por división de frecuencia* (OFDM), que difiere de los DMT en áreas de aplicación y en algunos aspectos de su diseño.

El OFDM se utiliza en la transmisión de datos por canales de transmisión de radio y canales de comunicación inalámbricos. Este dominio de aplicación requiere algunos cambios en

el diseño del sistema OFDM. A diferencia de los DMT que utilizan carga por la asignación de bits, el OFDM recurre a un número fijo de bits por subcanal. Esta restricción se vuelve necesaria porque un canal de transmisión implica transmisión de una vía, y en el ambiente de las comunicaciones inalámbricas el canal está variando demasiado rápido. Por tanto, en ambos casos no es factible para el transmisor conocer el canal y cómo "cargarlo".

Por tanto, el diagrama de bloques de la figura 6.60 se aplica igualmente a la OFDM, salvo cuando el codificador de la constelación de señales no incluye un algoritmo de carga para la asignación de bits. Además, otros dos cambios tienen que efectuarse al diseñar el sistema:

- En el transmisor, un *convertidor ascendente* se incluye después del convertidor digital-analógico para trasladar la frecuencia transmitida, facilitando de esa manera la propagación de la señal transmitida por el canal de radio.
- En el receptor se incluye un *convertidor descendente* antes del convertidor analógico-digital para deshacer la traslación de frecuencia que efectuó el ascendente en el transmisor.

Las aplicaciones del OFDM incluyen las siguientes:

1. *Comunicaciones inalámbricas.*
El OFDM, combinado con la codificación y el intercalamiento, ofrece una técnica efectiva para combatir el desvanecimiento de multitrayectoria, que constituye un rasgo característico de los canales de comunicación inalámbrica.

2. *Transmisión de audio digital.*
El OFDM se ha adoptado como el estándar para la transmisión de audio digital en Europa. En este caso el sistema implica también el uso combinado de la codificación y el intercalamiento.

(La codificación del control de errores y temas relacionados se presentan en el capítulo 10.)

6.14 *Sincronización*

La recepción coherente de una señal modulada digitalmente, sin que importe su forma, requiere que el receptor esté sincronizado con el transmisor. Afirmamos que dos secuencias de eventos (que representan a un transmisor y a un receptor) son *síncronas* entre sí cuando los eventos en una secuencia y los eventos correspondientes en la otra ocurren de manera simultánea. El proceso de hacer síncrona la situación, y de mantenerla en ese estado, recibe el nombre de *sincronización.*[23]

A partir de la discusión presentada acerca de la operación de técnicas de modulación digitales, reconocemos la necesidad de dos modos básicos de sincronización:

1. Cuando se usa detección coherente es necesario el conocimiento tanto de la frecuencia como de la fase portadora. La estimación de la fase y frecuencia de la portadora se denominan *recuperación de portadora* o *sincronización de portadora.*
2. Para efectuar la demodulación, el receptor tiene que conocer los instantes en los cuales la modulación puede cambiar su estado. Es decir, tiene que estar al tanto de los tiempos de inicio y de finalización de los símbolos individuales, de manera que pueda determinar cuándo muestrear y cuándo interrumpir los integradores de producto. La estimación de estos tiempos se conoce como *recuperación de reloj* o *sincronización de símbolos.*

Estos dos modos de sincronización pueden coincidir entre sí, o es posible que ocurran de modo secuencial uno luego del otro. De modo natural, en un sistema no coherente no interesa la sincronización de portadora.

La sincronización puede ponerse en práctica en una de dos maneras fundamentalmente diferentes:

1. *Sincronización asistida por datos.*
En los sistemas de sincronización asistida por datos se transmite un preámbulo junto con la señal que contiene los datos de una manera de multiplexado en el tiempo en una base periódica. El preámbulo contiene información acerca de la temporización de la portadora y del símbolo, la cual se extrae mediante el procesamiento apropiado de la salida del canal en el receptor. Un método de este tipo suele usarse en los satélites digitales y en las comunicaciones inalámbricas, donde el interés radica en minimizar el tiempo que se necesita para sincronizar el receptor con el transmisor. Presenta dos limitaciones: 1) Eficiencia reducida en el rendimiento de datos en la que se incurre al asignar cierta porción de la trama transmitida al preámbulo, y 2) eficiencia de potencia reducida al asignar cierta fracción de la potencia transmitida a la transmisión del preámbulo.

2. *Sincronización no asistida por datos.*
En este segundo método se evita el uso del preámbulo, y el receptor tiene la tarea de establecer la sincronización extrayendo la información necesaria de la señal modulada. Tanto el rendimiento como la eficiencia de potencia se mejoran de esa manera, aunque a costa de un aumento en el tiempo que se necesita para establecer la sincronización.

En cualquier caso, la sincronización constituye básicamente un problema de estimación de parámetros estadísticos. Un método principal para resolver un problema de este tipo es la *estimación de máxima verosimilitud* (vea la sección 5.5.), la cual se lleva a cabo formulando primero una función de verosimilitud logarítmica del parámetro de interés dada la señal recibida. Esta formulación es relativamente directa al tratar al ruido del canal como un proceso gaussiano. Más importante es el hecho de que no requiere información previa acerca de la señal modulada.

En esta sección centramos nuestra atención en las formas sin asistencia de datos de los sistemas de sincronización de portadora y temporización. En este contexto, es posible identificar dos procedimientos para resolver el problema de la sincronización, dada una señal modulada con portadora suprimida para conservar potencia:

1. *Método clásico.*
Un bloque constitutivo esencial en el método clásico para la sincronización es la *malla de fase encadenada.* (La malla de fase encadenada se presenta en el capítulo 2.) Específicamente, para la recuperación de la portadora el receptor requiere utilizar un *lazo de rastreo de portadora suprimida* para proporcionar una referencia de portadora (subportadora) secundaria coherente. Por ejemplo, podemos recurrir a una variante del lazo de Costas o al *lazo de potencia* M-ésima para la PSK M-aria. El lazo de Costas estándar para la modulación de doble banda lateral, portadora suprimida (DSB-SC) se explicó en el capítulo 2. En cuanto al lazo de potencia M-ésima, ésta consiste en la conexión en cascada de un dispositivo de ley de potencia M-ésima, un filtro pasobanda, una malla de fase encadenada y un divisor de frecuencia entre M. El objetivo aquí es explotar las propiedades de adquisición y rastreo de la malla de fase encadenada. Para una explicación adicional del lazo de potencia M-ésima, se refiere al lector al problema 6.47.

2. *Método algorítmico (moderno).*
En el método moderno, la solución a la estimación de máxima verosimilitud se formula en forma algorítmica utilizando el procesamiento de señales en tiempo discreto. Específicamente, la puesta en práctica del sincronizador se incorpora en un algoritmo que proporciona una estimación de la fase portadora o de la temporización del símbolo en un esquema de iteración por iteración. El procesamiento se efectúa en el dominio de la banda base para allanar el camino al uso del procesamiento de señales en tiempo discreto (digital).

En esta sección describimos el método algorítmico para la sincronización de sistemas PSK M-arios correspondientes tanto a la recuperación de la portadora como a la temporización del símbolo.

El método que se sigue en la exposición es secuencial porque la recuperación de la temporización se efectúa antes de la recuperación de fase. La razón de ello es que si conocemos el retardo de grupo en el que se incurre por la transmisión a través del canal, entonces una muestra por símbolo a la salida del filtro de acoplamiento en el receptor resulta suficiente para estimar la fase desconocida de la portadora. Además, la complejidad del cálculo del receptor se minimiza al utilizar algoritmos de sincronización que operan a la velocidad del símbolo $1/T$.

■ ALGORITMO RECURSIVO DE DECISIÓN DIRIGIDA PARA LA RECUPERACIÓN DE FASE

Como se subrayó antes, el primer paso importante para la solución del problema de sincronización consiste en formular la función de verosimilitud logarítmica correspondiente a la fase θ de la portadora, dada la señal recibida contaminada con ruido gaussiano. Sea $l(\theta)$ esta función de verosimilitud logarítmica, la cual sirve como función objetivo para estimar θ. El siguiente paso es determinar la derivada de $l(\theta)$ con respecto a θ. El paso final radica en formular un algoritmo (iterativo) para calcular paso por paso una verosimilitud de estimación máxima de la θ desconocida.

Evaluación de $\partial l(\theta)/\partial\theta$*

Sea $s_k(t)$ la señal transmitida para el símbolo $k = 0, 1, \ldots, M-1$:

$$s_k(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_c t + \alpha_k), \quad 0 \le t \le T \tag{6.238}$$

donde E es la energía del símbolo, T es el periodo del símbolo y

$$\alpha_k = 0, \frac{2\pi}{M}, \ldots, (M-1)\frac{2\pi}{M} \tag{6.239}$$

De manera equivalente, podemos escribir

$$s_k(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_c t + \alpha_k)g(t) \tag{6.240}$$

donde $g(t)$ es el pulso de formación, a saber, un pulso rectangular de amplitud unitaria y duración T. Sea τ_c el retardo de portadora (fase) y τ_g el retardo de envolvente (grupo), originados por el canal. Por definición, τ_c afecta a la portadora y τ_g a la envolvente. La señal recibida es entonces

$$\begin{aligned} x(t) &= \sqrt{\frac{2E}{T}}\cos(2\pi f_c(t-\tau_c)+\alpha_k)g(t-\tau_g)+w(t) \\ &= \sqrt{\frac{2E}{T}}\cos(2\pi f_c t+\theta+\alpha_k)g(t-\tau_g)+w(t) \end{aligned} \tag{6.241}$$

donde $w(t)$ es el ruido del canal y θ se define como $-2\pi f_c\tau_c$ para ser consistentes con la notación de la sección 6.6. Se desconocen tanto la fase θ de la portadora como el retardo de grupo τ_g. Sin embargo, se supone que permanecen constantes a lo largo del intervalo de observación $0 \le t \le T_0$ o a través de la transmisión de $L_0 = T_0/T$ símbolos. De modo equivalente, podemos escribir (utilizando τ en lugar de τ_g para simplificar las cosas)

$$x(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_c t+\theta+\alpha_k)+w(t), \qquad \tau \le t \le T+\tau \tag{6.242}$$

---

*El lector que no esté interesado en la deducción formal de $\partial l(\theta)/\partial\theta$ puede omitir esta sección y pasar a la siguiente subsección sin perder continuidad.

En el receptor las funciones básicas se definen por medio de

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t), \quad \tau \le t \le T + \tau \tag{6.243}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}} \operatorname{sen}(2\pi f_c t), \quad \tau \le t \le T + \tau \tag{6.244}$$

Aquí se supone que el receptor conoce perfectamente la frecuencia f_c de la portadora; en otro caso tiene que incluirse una compensación de frecuencia de portadora, lo cual complica el análisis. Por tanto, es posible representar la señal recibida $x(t)$ mediante el vector

$$\mathbf{x}(\tau) = \begin{bmatrix} x_1(\tau) \\ x_2(\tau) \end{bmatrix} \tag{6.245}$$

donde

$$x_i(\tau) = \int_{\tau}^{T+\tau} x(t)\phi_i(t)\, dt, \quad i = 1, 2 \tag{6.246}$$

De modo correspondiente, expresamos la componente de la señal de $\mathbf{x}(t)$ por medio del vector

$$\mathbf{s}(a_k, \theta, \tau) = \begin{bmatrix} s_1(a_k, \theta, \tau) \\ s_2(a_k, \theta, \tau) \end{bmatrix} \tag{6.247}$$

donde a_k es el símbolo transmitido y

$$s_i(a_k, \theta, \tau) = \int_{\tau}^{T+\tau} \sqrt{\frac{2E}{T}} \cos(2\pi f_c t + \theta + \alpha_k)\phi_i(t)\, dt \quad \text{para } i = 1, 2 \tag{6.248}$$

Suponiendo que f_c es un múltiplo entero de la velocidad del símbolo $1/T$, tenemos

$$s_1(a_k, \theta, \tau) = \sqrt{E} \cos(\theta + \alpha_k) \tag{6.249}$$

$$s_2(a_k, \theta, \tau) = \sqrt{E} \operatorname{sen}(\theta + \alpha_k) \tag{6.250}$$

En consecuencia, es posible escribir

$$\mathbf{x}_k(\tau) = \mathbf{s}(a_k, \theta, \tau) + \mathbf{w} \tag{6.251}$$

donde $\mathbf{w}$ es el vector de ruido

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{6.252}$$

con

$$w_i = \int_{\tau}^{T+\tau} w(t)\phi_i(t)\, dt, \quad i = 1, 2 \tag{6.253}$$

La w_i es el valor de muestreo de la variable aleatoria gaussiana de media cero y varianza $N_0/2$, donde $N_0/2$ es la densidad espectral de potencia (de los dos lados) del ruido del canal $w(t)$.

La función de densidad de probabilidad condicional del vector aleatorio $\mathbf{X}$, dada la transmisión del símbolo a_k y la ocurrencia de la fase θ de la portadora y del retardo de grupo τ, es

$$f_{\mathbf{X}}(\mathbf{x}|a_k, \theta, \tau) = \frac{1}{\pi N_0} \exp\left(-\frac{1}{N_0} \|\mathbf{x}_k(\tau) - \mathbf{s}(a_k, \theta, \tau)\|^2 \right) \tag{6.254}$$

para $a_k = 0$ la señal recibida $x(t)$ es igual al ruido del canal $w(t)$, por lo que

$$f_X(\mathbf{x}|a_k = 0) = \frac{1}{\pi N_0}\exp\left(-\frac{1}{N_0}\|\mathbf{x}_k(\tau)\|^2\right) \tag{6.255}$$

Por tanto, podemos definir la función de verosimilitud para la PSK M-aria en el receptor como

$$\begin{aligned} L(a_k,\theta,\tau) &= \frac{f_X(\mathbf{x}|a_k,\theta,\tau)}{f_X(\mathbf{x}|a_k = 0)} \\ &= \exp\left(\frac{2}{N_0}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k,\theta,\tau) - \frac{1}{N_0}\|\,\mathbf{s}(a_k,\theta,\tau)\,\|^2\right) \end{aligned} \tag{6.256}$$

En la PSK M-aria,

$$\|\mathbf{s}(a_k,\theta,\tau)\| = \text{constante}$$

cuando los puntos de mensaje se encuentran sobre un círculo de radio $\sqrt{E}$. De esa manera, ignorando el segundo término en el exponente, se puede simplificar la función de verosimilitud como

$$L(a_k,\theta,\tau) = \exp\left(\frac{2}{N_0}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k,\theta,\tau)\right) \tag{6.257}$$

Suponiendo que transmitimos una secuencia de L_0 símbolos independientes estadísticamente, a saber,

$$\mathbf{a} = [a_0, a_1, \ldots, a_{L_0-1}]^T \tag{6.258}$$

la función de verosimilitud resultante es

$$L(\mathbf{a},\theta,\tau) = \prod_{k=0}^{L_0-1}\exp\left(\frac{2}{N_0}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k,\theta,\tau)\right) \tag{6.259}$$

La función de verosimilitud logarítmica es entonces

$$\begin{aligned} l(a,\theta,\tau) &= \log L(a,\theta,\tau) \\ &= \frac{2}{N_0}\sum_{k=0}^{L_0-1}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k,\theta,\tau) \end{aligned} \tag{6.260}$$

A partir de las ecuaciones (6.249) y (6.250) deducimos

$$\begin{aligned} \hat{\mathbf{s}}_k(\theta) &= \mathbf{s}(\hat{a}_k,\theta,\tau) \\ &= \sqrt{E}\begin{bmatrix}\cos(\hat{\alpha}_k+\theta)\\ -\operatorname{sen}(\hat{\alpha}_k+\theta)\end{bmatrix}, \qquad k = 0, 1, \ldots, L_0-1 \end{aligned} \tag{6.261}$$

donde $\hat{\alpha}_k$ es una estimación de la α_k real que se produce en la salida del detector para el símbolo a_k. De modo correspondiente, podemos expresar la salida del filtro de acoplamiento como

$$\mathbf{x}_k = \begin{bmatrix} x_{1,k}\\ -x_{2,k}\end{bmatrix}$$

Por tanto, utilizando esta definición y la ecuación (6.261) en la ecuación (6.260), obtenemos

$$\begin{aligned} l(\theta) &= \frac{2\sqrt{E}}{N_0}\sum_{k=0}^{L_0-1}\left[x_{1,k}\cos(\hat{\alpha}_k+\theta) + x_{2,k}\operatorname{sen}(\hat{\alpha}_k+\theta)\right] \\ &= \frac{2\sqrt{E}}{N_0}\sum_{k=0}^{L_0-1}\left[(x_{1,k}\cos(\hat{\alpha}_k + x_{2,k}\operatorname{sen}\hat{\alpha}_k)\cos\theta\right. \\ &\qquad \left. -(x_{1,k}\operatorname{sen}\hat{\alpha}_k - x_{2,k}\cos\hat{\alpha}_k)\operatorname{sen}\theta\right] \end{aligned} \tag{6.262}$$

Diferenciando $l(\theta)$ con respecto a θ, obtenemos

$$\frac{\partial l(\theta)}{\partial \theta} = -\frac{2\sqrt{E}}{N_0} \sum_{k=0}^{L_0-1} \left[(x_{1,k} \cos \hat{\alpha}_k + x_{2,k} \operatorname{sen} \hat{\alpha}_k) \operatorname{sen} \theta + (x_{1,k} \operatorname{sen} \hat{\alpha}_k - x_{2,k} \cos \hat{\alpha}_k) \cos \theta \right] \tag{6.263}$$

Es posible simplificar la ecuación (6.263) introduciendo las siguientes notaciones:

$$\tilde{x}_k = x_{1,k} + jx_{2,k} \tag{6.264}$$

y

$$\begin{aligned} a_k &= e^{j\alpha_k} \\ &= \cos \alpha_k + j \operatorname{sen} \alpha_k \end{aligned} \tag{6.265}$$

donde $\tilde{x}_k$ es la envolvente compleja (es decir, el valor de banda base) de la salida del filtro de acoplamiento debida al k-ésimo símbolo transmitido y a_k es el indicador del símbolo en la constelación del mensaje de la PSK M-aria. De este modo podemos escribir

$$\begin{aligned} \operatorname{Re}[\hat{a}_k^* \tilde{x}_k] &= \operatorname{Re}[(\cos \hat{\alpha}_k - j \operatorname{sen} \hat{\alpha}_k)(x_{1,k} + jx_{2,k})] \\ &= x_{1,k} \cos \hat{\alpha}_k + x_{2,k} \operatorname{sen} \hat{\alpha}_k \end{aligned} \tag{6.266}$$

$$\begin{aligned} \operatorname{Im}[\hat{a}_k^* \tilde{x}_k] &= \operatorname{Im}[(\cos \hat{\alpha}_k - j \operatorname{sen} \hat{\alpha}_k)(x_{1,k} + jx_{2,k})] \\ &= -x_{1,k} \operatorname{sen} \hat{\alpha}_k + x_{2,k} \cos \hat{\alpha}_k \end{aligned} \tag{6.267}$$

También es posible advertir, de acuerdo con la fórmula de Euler:

$$e^{-j\theta} = \cos \theta - j \operatorname{sen} \theta \tag{6.268}$$

Por tanto, reescribimos la ecuación (6.263) en la forma compacta:

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \theta} &= \frac{2\sqrt{E}}{N_0} \sum_{k=0}^{L_0-1} \{(\operatorname{Re}[\hat{a}_k^* \tilde{x}_k])(\operatorname{Im}[e^{-j\theta}]) + (\operatorname{Im}[\hat{a}_k^* \tilde{x}_k])(\operatorname{Re}[e^{-j\theta}])\} \\ &= \frac{2\sqrt{E}}{N_0} \sum_{k=0}^{L_0-1} \operatorname{Im}[\hat{a}_k^* \tilde{x}_k e^{-j\theta}] \end{aligned} \tag{6.269}$$

donde $\hat{a}_k$ es una estimación de a_k, y el asterisco denota conjugación compleja.

ALGORITMO RECURSIVO PARA LA ESTIMACIÓN DE MÁXIMA VEROSIMILITUD DE LA FASE PORTADORA

Con la disposición de la fórmula de la ecuación (6.269) correspondiente a la derivada de la función de verosimilitud logarítmica $l(\theta)$ con respecto a la fase θ de la portadora, estamos listos para formular un algoritmo que busque maximizar $l(\theta)$. Nos gustaría realizar la maximización de manera iterativa a fin de que el receptor fuera capaz de responder a la señal recibida en una base de símbolo por símbolo. Con ese propósito, es posible formular la siguiente idea algorítmica tomada del filtrado adaptable (vea la discusión acerca del algoritmo de mínimos cuadrados que se presentó en los capítulos 3 y 4):

$$\begin{pmatrix} \text{Estimación} \\ \text{actualizada} \end{pmatrix} = \begin{pmatrix} \text{Estimación} \\ \text{anterior} \end{pmatrix} + \begin{pmatrix} \text{Parámetro del} \\ \text{tamaño del escalón} \end{pmatrix} \begin{pmatrix} \text{Señal de} \\ \text{error} \end{pmatrix} \tag{6.270}$$

donde la *señal de error*, o la señal de ajuste para ser más precisos, se define como el valor instantáneo del gradiente de la función de verosimilitud logarítmica $l(\theta)$ con respecto a θ. Note que el ajuste de parámetros que se aplica a la estimación anterior en la ecuación (6.270) es positivo

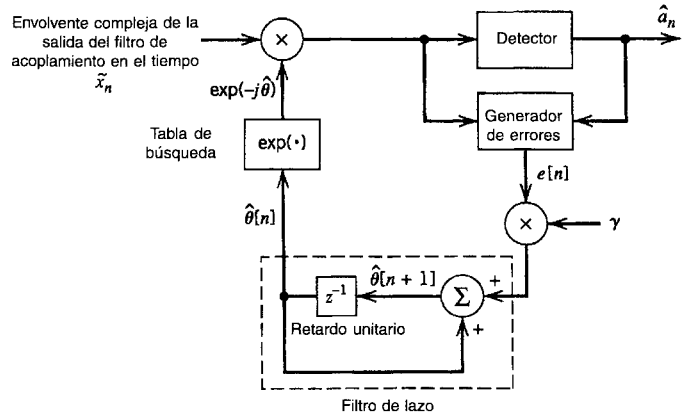


FIGURA 6.62 Lazo de Costas recursivo.

cuando el objetivo aquí es efectuar el *ascenso de gradiente*. En la ecuación (6.269) vemos de inmediato que la señal de error (es decir, el valor instantáneo de $\partial l(\theta)/\partial\theta$ debido a la transmisión de un solo símbolo) está dada por

$$e[n] = \text{Im}\left[\hat{a}_n^* \tilde{x}_n e^{-j\hat{\theta}}\right] \tag{6.271}$$

donde el factor de escalamiento $2\sqrt{E}/N_0$ se explica a continuación. Además hemos utilizado n en lugar de k para denotar el periodo o iteración del algoritmo. Por tanto, empleamos la ecuación (6.270) para escribir

$$\hat{\theta}[n+1] = \hat{\theta}[n] + \gamma e[n] \tag{6.272}$$

donde $\hat{\theta}[n]$ es la *estimación anterior* de la fase θ de la portadora, $\hat{\theta}[n+1]$ es la *estimación actualizada* de θ, y γ es el *parámetro del tamaño de escalón*; el factor de escalamiento $2\sqrt{E}/N_0$ se absorbe en γ.

Las ecuaciones (6.271) y (6.272) definen el algoritmo recursivo para la recuperación de fase. Este mismo se implementa utilizando el sistema que se presenta en la figura 6.62, el cual puede considerarse como una generalización recursiva del lazo de Costas. Por tanto, podemos referirnos a este como el *lazo de Costas recursivo* para la sincronización de fase.

Los siguientes puntos deben advertirse en la figura 6.62:

- El detector ofrece una estimación del símbolo transmitido a_n, dada la salida del filtro de acoplamiento.
- La tabla de búsqueda proporciona el valor de $\exp(-j\hat{\theta}[n]) = \cos\hat{\theta}[n] - \text{sen}\,\hat{\theta}[n]$ para una entrada $\hat{\theta}[n]$.
- La salida del generador de errores es la señal de error en $e[n]$.
- El bloque marcado z^{-1} es un elemento de retardo unitario con el retardo igual al periodo del símbolo T.

El lazo de Costas recursivo de la figura 6.62 utiliza un filtro digital de primer orden. Para mejorar el desempeño del rastreo de este sistema de sincronización es posible utilizar un filtro digital de segundo orden. La figura 6.63 muestra un ejemplo de un filtro digital de segundo orden conformado por una cascada de dos secciones de primer orden con ρ como un *parámetro de lazo* ajustable. Una importante propiedad de un filtro de segundo orden utilizado en el lazo de Costas para la recuperación de fase es que a la larga se cerrará sobre la portadora entrante sin error estático, siempre y cuando el error de la frecuencia entre el receptor y el transmisor sea pequeño inicialmente.

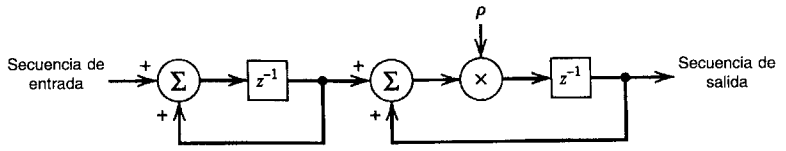


FIGURA 6.63 Filtro digital de segundo orden.

ALGORITMO RECURSIVO SIN ASISTENCIA DE DATOS PARA LA TEMPORIZACIÓN DEL SÍMBOLO

Para la sincronización de la temporización la única suposición que se realiza es que el receptor conoce la frecuencia f_c de la portadora. El requerimiento estriba en formular un algoritmo para la estimación recursiva del retardo de grupo τ en el que se incurre en el curso de la transmisión de la señal modulada a través del canal.

Sea $L(a_k, \theta, \tau)$ la función de verosimilitud de τ, que es también una función del símbolo transmitido a_k y la fase θ de la portadora. La función de verosimilitud se define mediante la ecuación (6.257). Para seguir adelante debemos eliminar las dependencias de $L(a_k, \theta, \tau)$ de la secuencia de datos transmitidos $\{a_k\}$ y de la fase θ de la portadora, como se describe a continuación.

Para eliminar la dependencia de θ promediamos la función de verosimilitud $L(a_k, \theta, \tau)$, pero *no* su logaritmo sobre todos los valores posibles de θ dentro del intervalo $[0, 2\pi]$. Suponiendo que θ se distribuye de manera uniforme dentro de este intervalo, lo cual suele ser justificable, es posible escribir

$$\begin{aligned} L_{\text{pro}}(a_k, \tau) &= \int_0^{2\pi} L(a_k, \theta, \tau) f_\Theta(\theta)\, d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \exp\left(\frac{2}{N_0}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k, \theta, \tau)\right) d\theta \end{aligned}$$

El exponente en $L(a_k, \theta, \tau)$ se expresa mediante (vea el problema 6.49)

$$\begin{aligned} \frac{2}{N_0}\mathbf{x}_k^T(\tau)\mathbf{s}(a_k, \theta, \tau) &= \frac{2\sqrt{E}}{N_0}\text{Re}\left[a_k^* \tilde{x}_k(\tau) e^{-j\theta}\right] \\ &= \frac{2\sqrt{E}}{N_0}\text{Re}\left[\left|a_k \tilde{x}_k(\tau)\right| \exp(j(\arg[\tilde{x}_k(\tau)] - \arg[a_k] - \theta))\right] \\ &= \frac{2\sqrt{E}}{N_0}\left|a_k \tilde{x}_k(\tau)\right| \cos(\arg[\tilde{x}_k(\tau)] - \arg[a_k] - \theta) \end{aligned} \tag{6.273}$$

Por lo tanto,

$$\begin{aligned} L_{\text{pro}}(a_k, \tau) &= \frac{1}{2\pi}\int_0^{2\pi} \exp\left(\frac{2\sqrt{E}}{N_0}\left|a_k \tilde{x}_k(\tau)\right| \cos(\arg[\tilde{x}_k(\tau)] - \arg[a_k] - \theta)\right) d\theta \\ &= \frac{1}{2\pi}\int_{-\arg[\tilde{x}_k(\tau)]+\arg[a_k]}^{2\pi-\arg[\tilde{x}_k(\tau)]+\arg[a_k]} \exp\left(\frac{2\sqrt{E}}{N_0}\left|a_k \tilde{x}_k(\tau)\right| \cos(\varphi)\, d\varphi\right) \end{aligned} \tag{6.274}$$

donde, en la última línea, hemos efectuado la sustitución

$$\varphi = \arg[\tilde{x}_k(\tau)] - \arg[a_k] - \theta$$

Apelamos ahora a la definición de la función de Bessel modificada de orden cero, de acuerdo a como indica (vea el apéndice 3)

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos\varphi}\, d\varphi \tag{6.275}$$

De ese modo podemos expresar la función de verosimilitud promedio $L_{\text{pro}}(a_k, \tau)$ como

$$L_{\text{pro}}(a_k, \tau) = I_0\left(\frac{2\sqrt{E}}{N_0}\left|a_k \tilde{x}_k(\tau)\right|\right) \tag{6.276}$$

donde $\tilde{x}_k(\tau)$ es la envolvente compleja de la salida del filtro de acoplamiento en el receptor debida al k-ésimo símbolo transmitido a_k. Para la PSK M-aria tenemos

$$|a_k| = 1 \quad \text{para toda } k$$

De esa manera, la ecuación (6.276) se reduce a

$$L_{\text{pro}}(a_k, \tau) = I_0\left(\frac{2\sqrt{E}}{N_0}\left|\tilde{x}_k(\tau)\right|\right) \tag{6.277}$$

De este modo vemos que promediar la función de verosimilitud sobre la fase θ de la portadora ha eliminado también la dependencia del símbolo transmitido a_k para la PSK M-aria.

Por último, teniendo en cuenta la transmisión de L_0 símbolos independientes $a_0, a_1, \ldots, a_{L_0-1}$, podemos expresar la función de verosimilitud total

$$\begin{aligned} L_{\text{pro}}(\tau) &= \prod_{k=0}^{L_0-1} L_{\text{pro}}(a_k, \tau) \\ &= \prod_{k=0}^{L_0-1} I_0\left(\frac{2\sqrt{E}}{N_0}\left|\tilde{x}_k(\tau)\right|\right) \end{aligned} \tag{6.278}$$

Después podemos tomar el logaritmo de $L_{\text{pro}}(\tau)$ para obtener la función de verosimilitud logarítmica de τ como

$$\begin{aligned} l_{\text{pro}}(\tau) &= \log L_{\text{pro}}(\tau) \\ &= \sum_{k=0}^{L_0-1} \log I_0\left(\frac{2\sqrt{E}}{N_0}\left|\tilde{x}_k(\tau)\right|\right) \end{aligned} \tag{6.279}$$

Para seguir más adelante, necesitamos aproximar $l_{\text{pro}}(\tau)$. Con ese propósito advertimos primero que la función de $I_0(x)$ puede expandirse en una serie de potencias como (vea el apéndice 3)

$$I_0(x) = \sum_{m=0}^{\infty} \frac{\left(\frac{1}{2}x\right)^{2m}}{(m!)^2}$$

Para valores pequeños de x podemos aproximar entonces $I_0(x)$ como

$$I_0(x) \simeq 1 + \frac{x^2}{4}$$

Es posible simplificar aún más las cosas utilizando la aproximación

$$\begin{aligned} \log I_0(x) &\simeq \log\left(1 + \frac{x^2}{4}\right) \\ &\simeq \frac{x^2}{4} \quad \text{para } x \text{ pequeña} \end{aligned}$$

Para el problema presente, x pequeña corresponde a una relación señal a ruido pequeña. De acuerdo con esta condición, es viable aproximar la ecuación (6.279) como

$$l_{\text{pro}}(\tau) \simeq \frac{E}{N_0^2} \sum_{k=0}^{L_0-1} \left|\tilde{x}_k(\tau)\right|^2 \tag{6.280}$$

donde como mencionamos antes, $\tilde{x}_k(\tau)$ es la envolvente compleja de la salida del filtro de acoplamiento debida al k-ésimo símbolo transmitido.

Al diferenciar $l_{\text{pro}}(\tau)$ con respecto al retardo de grupo τ, obtenemos

$$\begin{aligned}\frac{\partial l_{\text{pro}}(\tau)}{\partial \tau} &= \frac{E}{N_0^2}\sum_{k=0}^{L_0-1}\frac{\partial}{\partial \tau}\left|\tilde{x}_k(\tau)\right|^2 \\ &= \frac{2E}{N_0^2}\sum_{k=0}^{L_0-1}\text{Re}[\tilde{x}_k^*(\tau)\tilde{x}_k'(\tau)\tilde{x}_k'(\tau)]\end{aligned} \tag{6.281}$$

donde $\tilde{x}_k^*(\tau)$ es la conjugada compleja de $\tilde{x}_k(\tau)$ y $\tilde{x}_k'(\tau)$ es su derivada con respecto a τ. Por tanto, es posible definir la señal de error para la recuperación de la temporización como (explicando más adelante el factor de escala $2E/N_0^2$)

$$e[n] = \text{Re}[\hat{x}_n^*(\tau)\hat{x}_n'(\tau)]$$

donde hemos utilizado n en lugar de k para ser consistentes con la notación de la figura 6.62. Sea $\hat{\tau}_n$ la estimación del retardo desconocido τ en el tiempo $t = nT$. Entonces, introduciendo las definiciones

$$\tilde{x}_n(\tau) = \tilde{x}(nT + \hat{\tau}_n)$$

y

$$\tilde{x}_n'(\tau) = \tilde{x}'(nT + \hat{\tau}_n)$$

podemos reformular la señal de error $e[n]$ como

$$e[n] = \text{Re}[\tilde{x}^*(nT + \hat{\tau}_n)\tilde{x}'(nT + \hat{\tau}_n)] \tag{6.282}$$

El cálculo de la señal de error requiere el uso de dos filtros:

1. El filtro de acoplamiento complejo para la generación $\tilde{x}_n(\tau)$.
2. El filtro de acoplamiento derivativo para generar $\tilde{x}_n'(\tau)$.

El receptor ya está equipado con el primer filtro. El segundo es nuevo. En la práctica, la complejidad computacional agregada por causa del filtro de acoplamiento derivativo resulta objetable. Es posible eliminarlo utilizando una diferencia finita para aproximar la derivada $\tilde{x}_n'(\tau)$ como

$$\tilde{x}'(nT + \hat{\tau}_n) \simeq \frac{1}{T}\left[\tilde{x}\left(nT + \frac{T}{2} + \hat{\tau}_{n+1/2}\right) - \tilde{x}\left(nT - \frac{T}{2} + \hat{\tau}_{n-1/2}\right)\right] \tag{6.283}$$

donde $\hat{\tau}_{n\pm 1/2}$ son las estimaciones de la temporización calculadas en $nT \pm T/2$. Resulta deseable efectuar una modificación adicional para explicar el hecho de que las estimaciones de la temporización se actualizan en múltiplos del periodo del símbolo T y las únicas cantidades disponibles son las $\hat{\tau}_n$. En consecuencia, sustituimos $\hat{\tau}_{n+1/2}$ por $\hat{\tau}_n$ (la cual representa la estimación más reciente de τ) y $\hat{\tau}_{n-1/2}$ por $\hat{\tau}_{n-1}$ (que es la estimación de τ antes de la última). Podemos reescribir de ese modo la ecuación (6.283) como

$$\tilde{x}'(nT + \hat{\tau}_n) \quad \frac{1}{T}\left[\tilde{x}\left(nT + \frac{T}{2} + \hat{\tau}_n\right) - \tilde{x}\left(nT - \frac{T}{2} + \hat{\tau}_{n-1}\right)\right] \tag{6.284}$$

y, de esa manera, redefinir por último la señal de error como

$$e[n] = \text{Re}\left\{\tilde{x}^*(nT + \hat{\tau}_n)\left[\tilde{x}\left(nT + \frac{T}{2} + \hat{\tau}_n\right) - \tilde{x}\left(nT - \frac{T}{2} + \hat{\tau}_{n-1}\right)\right]\right\} \tag{6.285}$$

de la cual el factor de escalamiento $1/T$ se explica más adelante.

Estamos listos ahora para formular el algoritmo recursivo para la recuperación de la temporización:

$$c[n+1] = c[n] + \gamma e[n] \tag{6.286}$$

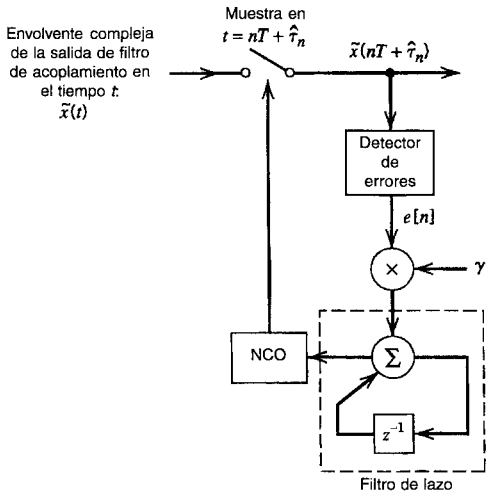


FIGURA 6.64 Sincronizador de retardo temprano-tardío sin asistencia de datos.

donde γ es el parámetro del tamaño del escalón en el cual se absorben $2E/N_0^2$ y $1/T$, y la señal de error $e[n]$ se define mediante la ecuación (6.285). La $c[n]$ es un número real que se emplea como el control para la frecuencia de un oscilador, conocida como un *oscilador controlado por número* (NCO). El esquema para poner en práctica el algoritmo de recuperación de la temporización de las ecuaciones (6.285) y (6.286) se muestra en la figura 6.64. Este esquema es análogo a la versión en tiempo continuo del sincronizador de compuerta temprano-tardía que se usa ampliamente en la recuperación de la temporización. Por tanto, éste recibe el nombre de *sincronizador de retardo temprano-tardío sin asistencia de datos (NDA-ELD)*. En cada iteración, éste trabaja en tres muestras sucesivas de la salida del filtro de acoplamiento, a saber,

$$\tilde{x}\left(nT + \frac{T}{2} + \hat{\tau}_n\right),\ \tilde{x}(nT + \hat{\tau}_n) \text{ y } \tilde{x}\left(nT + \frac{T}{2} - \hat{\tau}_{n-1}\right)$$

La primera muestra es la temprana y la última la tardía, ambas con base en la intermedia.

Advierta que podíamos haber simplificado las derivaciones que se presentaban en esta sección al utilizar la transformación de pasobanda a pasobaja compleja que se describe en el apéndice 2. No lo hicimos sólo por simplificar la comprensión del material que se presenta aquí.

6.15 *Experimentos de computadora: recuperación de portadora y temporización del símbolo*

En esta sección ilustramos la operación de lazo de Costas recursivo y del sincronizador temprano-tardío al considerar un sistema QPSK coherente con las siguientes dos especificaciones:

i) Respuesta del canal: coseno elevado (Nyquist) con factor de caída $\alpha = 0.5$.
ii) Filtro de lazo: filtro digital de primer orden con su función de transferencia definida por

$$H(z) = \frac{1}{z - (1 - \gamma A)} \tag{6.287}$$

donde γ es el parámetro del tamaño de escalón y A es un parámetro por definir.
iii) Ancho de banda del lazo, $B_L = 2\%$ de la velocidad del símbolo $1/T$, es decir, $B_L T = 0.02$.

Experimento 1: recuperación de fase portadora

Para investigar el comportamiento de la adquisición de fase del lazo de Costas recursivo, necesitamos tener la llamada *curva* S del generador de errores de fase. Ésta se define como la esperanza de la señal de ajuste en $e[n]$ condicionada a un valor fijo de error de fase

$$\varphi = \theta - \hat{\theta}$$

donde θ es el valor real de la fase portadora y $\hat{\theta}$ es su estimación. Esto es,

$$S(\varphi) = E[e[n]|\varphi] \tag{6.288}$$

Experimentalmente, $S(\varphi)$ se mide abriendo el lazo de Costas recursivo de la figura 6.62 y midiendo el promedio de la señal de ajuste $e[n]$, como se indica en la figura 6.65.

El procedimiento de puesta en práctica consiste en los siguientes pasos. Primero se genera la envolvente compleja de la señal recibida, la cual está dada por

$$\tilde{x}_k(t) = \sqrt{\frac{2E}{T}} \exp(-j(2\pi f_c \tau_c - \alpha_k)) g(t - \tau_g) + \tilde{w}(t) \tag{6.289}$$

donde $\alpha_k = 0, \pi/2, \pi, 3\pi/4$; τ_c es el retardo de la portadora y τ_g el retardo de grupo, $\tilde{w}(t)$ es el ruido del canal de valor complejo. La respuesta total del canal $g(t)$ está dada por el pulso de Nyquist (vea la sección 4.5)

$$g(t) = \frac{\text{sen}(\pi t/T)}{(\pi t/T)} \cdot \frac{\cos(\pi \alpha t/T)}{1 - 4\alpha^2 t^2/T^2} \tag{6.290}$$

donde $\alpha = 0.5$. Como se señaló antes, se supone que se conoce la temporización del símbolo (es decir, el retardo de grupo τ_g), y el problema estriba en estimar la fase portadora $\theta = -2\pi f_c \tau_c$. El efecto de θ es correr un elemento en la constelación de señales de la manera en que se indica en la figura 6.66.

Empleando el procedimiento experimental que se describe en la figura 6.65, la curva S del sistema QPSK puede medirse ahora. La figura 6.67*a* muestra la curva S ideal, suponiendo una relación señal a ruido infinitamente grande. Esta curva presenta discontinuidades en $\varphi = \pm m\pi/4$, donde $m = 0, 1, 3, \ldots$, debido a la ambigüedad que se encuentra en la detección del símbolo

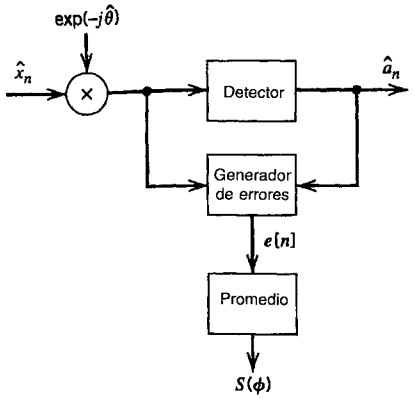


FIGURA 6.65 Esquema para medir la curva S para la recuperación de fase portadora.

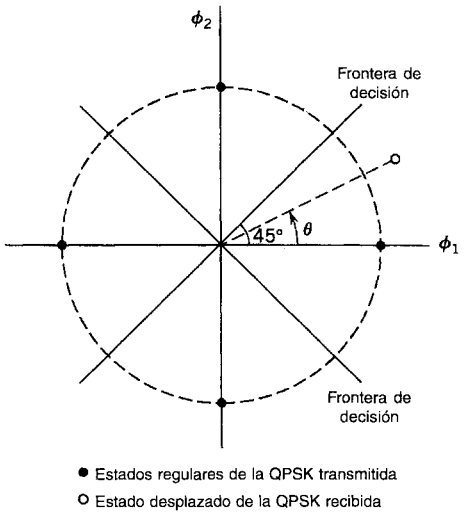


FIGURA 6.66 Ilustración del efecto de la fase θ de la portadora en un estado de la señal QPSK.

transmitido a_k. La presencia de ruido del canal tiende a redondear las discontinuidades, como se muestra en la curva S medida experimentalmente de la figura 6.67*b*. Los resultados que se presentan en esta última figura se obtuvieron para $E/N_0 = 10$ dB. Recuerde que las componentes en fase y en cuadratura del ruido de banda angosta tienen una distribución gaussiana idéntica con media cero y la misma varianza que la del ruido de banda angosta original; estas dos componentes definen a $\tilde{w}(t)$.

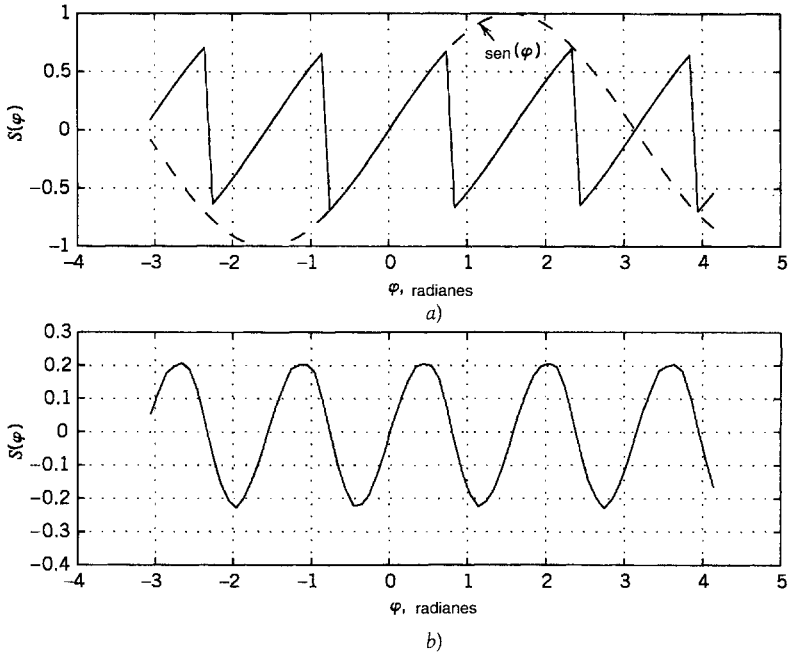


FIGURA 6.67 Desempeño del lazo de Costas recursivo. *a*) Curva S para $(E/N_0) = \infty$. *b*) Curva S para $(E/N_0) = 10$ dB.

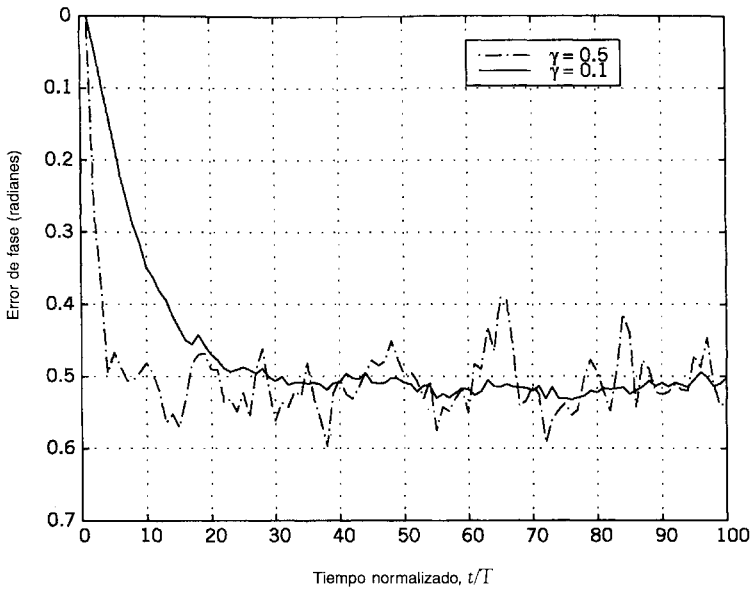


FIGURA 6.68 Efectos de la variación del parámetro del tamaño del escalón sobre el comportamiento relativo a la convergencia del lazo de Costas recursivo.

Cuando se han establecido las condiciones en estado estable, la fase estimada $\hat{\theta}$ fluctuará alrededor del valor real θ. El grado de estas fluctuaciones depende del parámetro de escalón γ y de la relación señal recibida a ruido:

i) En la figura 6.68 se grafica el error de fase φ en función del tiempo normalizado t/T para dos valores diferentes del parámetro del tamaño de escalón γ, a saber, 0.1 y 0.5, y $E/N_0 = 20$ dB fija. Esta figura muestra claramente que tanto más pequeña hacemos la γ tanto menor resultarán las fluctuaciones en estado estable del error de fase φ. Sin embargo, esta mejora se consigue a costa de una velocidad inferior de la convergencia del algoritmo. El número de iteraciones necesarias para que el algoritmo alcance el estado estable está dado de manera aproximada por

$$L_0 \simeq \frac{1}{2B_L T} \tag{6.291}$$

El propio ancho de banda normalizado $B_L T$ está dado aproximadamente por

$$B_L T \simeq \frac{\gamma A}{4} \tag{6.292}$$

donde A es la pendiente de la curva S en el origen. Para $\gamma = 0.1$ y $B_L T = 0.02$, la ecuación (6.291) produce $L_0 = 25$ iteraciones, las cuales coinciden con la curva continua que se grafica en la figura 6.68. Además, de las ecuaciones (6.291) y (6.292) vemos que L_0 es inversamente proporcional a γ, lo cual coincide también con los resultados que se presentan en la figura 6.68.

ii) En la figura 6.69 se grafica el error de fase φ en función del tiempo normalizado t/T para tres valores diferentes de E/N_0, a saber, 5, 10 y 30 dB, y $\gamma = 0.08$ fija. En este caso vemos que tanto más grande hacemos la relación señal a ruido, tanto menor serán las fluctuaciones en estado estable en el error de fase φ. Además, la tasa de convergencia del algoritmo también mejora con el aumento de la relación señal a ruido, lo cual se satisface de manera intuitiva.

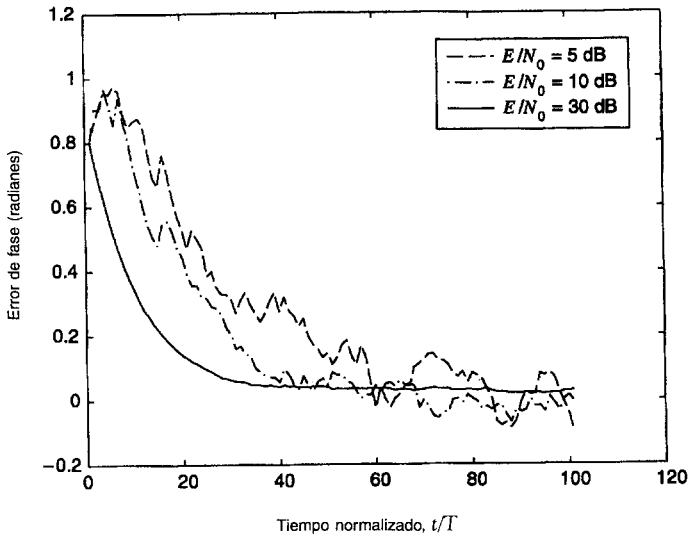


FIGURA 6.69 Comportamiento relativo a la convergencia del lazo de Costas recursivo para E/N_0 variable.

La figura 6.70 grafica la varianza de error de fase (promediado sobre 100 ensayos del experimento) en función de E/N_0 (medida en decibeles) para $B_L T = 0.02$ y $\gamma = 0.08$. Esta figura incluye también una gráfica de la *cota de Cramér-Rao modificada* definida por[24]

$$\text{MCRB}(\theta) = \frac{1}{2L_0(E/N_0)} \tag{6.293}$$

Esta cota es una modificación de la cota ordinaria de Cramér-Rao, que es un límite inferior sobre la varianza de cualquier estimador *no polarizado*. La modificación de esta cota se realiza para superar las dificultades de cómputo que se encuentran en los problemas de sincronización prácticos.

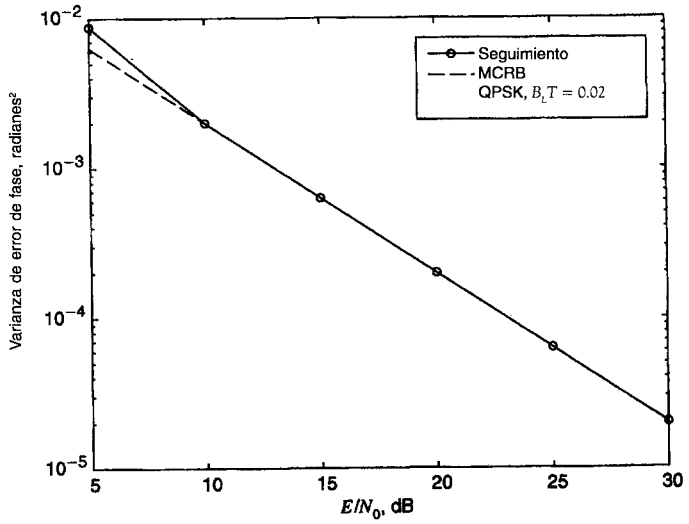


FIGURA 6.70 Comparación de la varianza de error de seguimiento medido del lazo de Costas recursivo contra la teoría para E/N_0 variable.

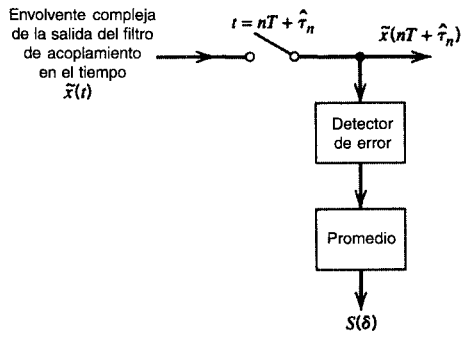


FIGURA 6.71 Esquema para medir la curva S para el sincronizador de retardo temprano-tardío recursivo.

En cualquier caso, los resultados experimentales y teóricos que se presentan en la figura 6.70 concuerdan mucho para $(E/N_0) \geq 10$ dB.

Experimento 2: recuperación de la temporización del símbolo

Para medir la curva S relativa al sincronizador de retardo temprano-tardío sin asistencia de datos para la recuperación de la temporización del símbolo, es posible utilizar el arreglo experimental que se muestra en la figura 6.71, donde la δ en $S(\delta)$ se refiere a la compensación de la temporización. La curva S medida en ese modo se grafica para la figura 6.72 para $E/N_0 = 10$ dB y $E/N_0 = \infty$.

La figura 6.73 grafica el valor normalizado del error medido experimentalmente de la temporización del símbolo en función de E/N_0 para dos valores diferentes del parámetro de escalón γ, a saber, $T/20$ y $T/200$. Esta figura incluye también gráficas teóricas de la correspondiente cota modificada de Cramér-Rao de la ecuación (6.293) adaptada al error de temporización del símbolo. A partir de los resultados que se presentan aquí, observamos que cuando se reduce el parámetro del

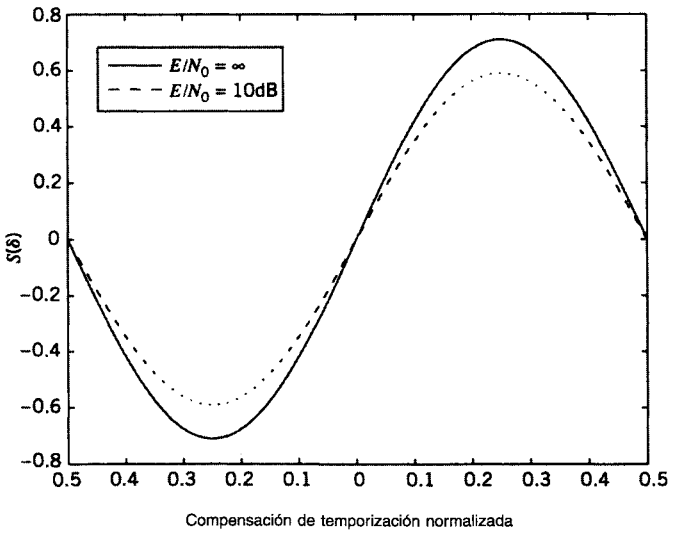


FIGURA 6.72 Curva S del sincronizador de retardo temprano-tardío sin asistencia de datos medida en condiciones sin y con ruido.

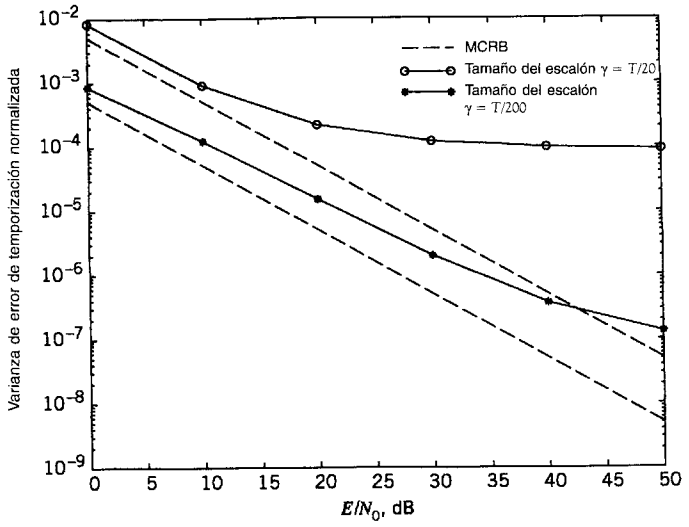


FIGURA 6.73 Comparación de la varianza de error de seguimiento del sincronizador de retardo temprano-tardío sin asistencia de datos contra la teoría para E/N_0 variable y dos parámetros de tamaño de escalón.

tamaño del escalón γ, el error de temporización normalizado se reduce y se alarga el intervalo de E/N_0 para el cual es válida la cota modificada de Cramér-Rao (aunque de un modo aproximado).

6.16 *Resumen y discusión*

Con la teoría de fundamentos básicos de los receptores óptimos del capítulo 5 a nuestra disposición, en este capítulo obtuvimos fórmulas para, o cotas sobre, la tasa de error de bits de algunas técnicas importantes de modulación digital en un canal AWGN:

1. Corrimiento de fase por llaveo (PSK), representado por
   - Corrimiento de fase por llaveo binaria (BPSK).
   - Corrimiento de cuadrifase por llaveo coherente (QPSK) y sus variantes, a saber, la QPSK de compensación y la QPSK desfasada $\pi/4$
   - La PSK M-aria coherente, que incluye la BPSK y la QPSK como casos especiales con M = 2 y M = 4, respectivamente. La PSK M-aria coherente se usa en las comunicaciones digitales por satélite.
   - Corrimiento de fase por llaveo diferencial (DPSK), que puede considerarse como una forma pseudo-no coherente de la PSK.
2. La modulación de amplitud en cuadratura M-aria coherente (QAM), que es una forma híbrida de modulación que combina el corrimiento en amplitud y en fase por llaveo. Para M = 4 incluye a la QPSK como un caso especial. La QAM M-aria es básica para la construcción de módems de alta velocidad.
3. Corrimiento de frecuencia por llaveo (FSK), representada por
   - Corrimiento de frecuencia por llaveo binaria coherente.
   - Formas coherentes de corrimiento mínimo por llaveo (MSK) y corrimiento mínimo por llaveo gaussiana GMSK, que es básica para la construcción de comunicaciones inalámbricas GSM.

- La FSK M-aria coherente.
- La FSK binaria no coherente.

En este capítulo estudiamos también dos técnicas alternativas para la transmisión de datos pasobanda: modulación de amplitud/fase sin portadora (CAP) y multitonos discretos (DMT). En el caso de un canal AWGN, el desempeño de la CAP y de los DMT es equivalente, debido a que estos últimos pueden considerarse como una transformación reversible lineal de una señal modulada de una sola portadora. Sin embargo, se desempeñan de modo bastante diferente en la práctica y se apartan de este modelo idealizado.[25] Los DMT se han estandarizado para utilizarse en líneas digitales asimétricas de suscriptor (ADSL) utilizando pares trenzados. La CAP, con el uso de ecualización de decisión realimentada, ofrece otro método para resolver el problema de las ADSL. La CAP se usa también para la transmisión de datos en redes de área local en sistemas de distribución de instalaciones.

Los DMT constituyen una forma de modulación multicanal y por ello corresponden a un multiplexado por división de frecuencia ortogonal (OFDM). La diferencia básica entre los DMT y el OFDM es que los primeros permiten el uso de carga para optimizar la capacidad de información, en tanto que no ocurre así con el OFDM. Esta diferencia surge debido a sus diferentes dominios de aplicación. Los DMT se aplican a canales de dos alambres, como las ADSL, en tanto que el OFDM se emplea en la transmisión y en los canales inalámbricos.

Independientemente del sistema de modulación digital de interés, la sincronización del receptor y el transmisor resulta esencial para la operación del sistema. Se requiere la recuperación de la temporización del símbolo ya sea si el receptor es coherente o no. Si el receptor es coherente, requerimos también la provisión para la recuperación de la portadora. En la última parte del capítulo explicamos los sincronizadores sin asistencia de datos para proveer estos dos requerimientos destacando el corrimiento de fase por llaveo M-aria de señales en las cuales se suprime la portadora. La presentación se enfocó en las técnicas de sincronización iterativas que son adecuadas de modo natural en el procesamiento de señales digitales.

NOTAS Y BIBLIOGRAFÍA

1. En lo que se refiere a un artículo tutorial en el que se repasan diferentes técnicas de modulación digital (ASK, FSK y PSK) con base en un punto de vista geométrico, vea Arthurs y Dym [189]. Consulte también la siguiente lista de libros:
   - Anderson (capítulo 3)
   - Benedetto y Biglieri ([11], capítulos 4 y 5)
   - Lee y Messerschmitt ([85], parte II)
   - Proakis ([123], capítulo 5)
   - Sklar ([140], capítulo 3)
   - Viterbi y Omura ([162], pp. 47-127)
2. Con respecto a uno de los primeros artículos sobre la QPSK de compensación, vea Gitlin y Ho [262].
3. Akaiwa y Nagata [182] describieron por primera vez en una publicación a la QPSK desfasada $\pi/4$.
4. Chennakeshu y Sauliner [218] usaron simulaciones de computadora para estudiar el desempeño de la QPSK con corrimiento de $\pi/4$ en un ambiente de comunicaciones inalámbricas digitales. La señal de formación de pulso que se utiliza en la generación de la señal QPSK desfasada $\pi/4$ se basa en el espectro de la raíz cuadrada del coseno elevado (vea el problema 4.38). En este último artículo se demuestra que el desempeño de la QPSK desfasada $\pi/4$ puede degradarse rápidamente en un ambiente de estas características. El detector diferencial de la figura 6.13 se basó en Chennakeshu y Sauliner [218].

5. Para una deducción de la ecuación (6.65), vea Cioffi [30].

6. La deducción de la ecuación (6.76) fue informada por primera vez en 1975 en un memorándum interno de Bell Laboratories firmado por Werner. Poco después Falconer [242] emitió otro memorándum de dichos laboratorios, en el cual apuntaba que la rotación del símbolo no era realmente necesaria si no se deseaba la compatibilidad con la QAM existente u otras señales pasobanda, lo que simplificaba la representación matemática de las señales CAP (y, por tanto, su puesta en práctica), como se indica en la ecuación (6.77). Sin embargo, la terminología "CAP" no se acuñó hasta 1987 cuando la modulación de amplitud/fase sin portadora fue sustituida por la CAP mediante los estándares representativos, Garry Smith, de Bell Laboratories. La primera explicación detallada de la CAP en el contexto de las líneas digitales de suscriptor se presentó en un informe de dos partes de Werner [402, 403]. En un informe separado de Chen, Im y Werner [217] se estudiaba la factibilidad de la CAP para su empleo en las líneas digitales de suscriptor. Consulte también el libro de Chen [29], pp. 461-473. La aplicación de la CAP en las redes de área local, y que lleva el uso de pares trenzados en longitudes menores que 100 m, se analiza en el artículo de Im y Werner [282]; la longitud máxima de 100 m se especifica por medio de un estándar para el alambrado de instalaciones.

   La puesta en práctica digital de un ecualizador de banda base similar al receptor CAP de la figura 6.24 se explica en Mueller y Werner [326].

7. La señal MSK la describieron por primera vez Doelz y Heald [236]. En relación con una revisión instructiva de la MSK y en comparación con la QPSK, vea Pasupathy [339]. Puesto que el espaciamiento de frecuencia es sólo la mitad del espaciamiento convencional de $1/T_b$ que se utiliza en la detección coherente de señales FSK binarias, este esquema de transmisión de señales también se conoce como la FSK *rápida*. Vea deBuda [229].

8. En lo que se refiere a las primeras explicaciones de la MSK gaussiana, vea Murota y Hirade [327] e Ishizuka y Hirade [284].

9. La especificación analítica de la densidad espectral de potencia de la FM digital resulta difícil de manejar, salvo en el caso de un pulso modulador de forma rectangular. Un artículo de Garrison presenta un procedimiento que se basa en la selección de una aproximación apropiada de la duración limitada/nivel de cuantización para el pulso modulador. Las ecuaciones que se formulan ahí son en particular adecuadas para los cálculos de computadora de los efectos de potencia de señales de FM digitales; consulte el libro de Stüber.

10. Un análisis detallado de los espectros de la FSK M-aria para un valor arbitrario de desviación de frecuencia se presenta en el artículo de Anderson y Salz [186]. Los resultados que se muestran graficados en la figura 6.36 representan un caso especial de una fórmula obtenida en ese artículo para una desviación de frecuencia para $k = 0.5$.

11. El método estándar de deducción de la tasa de error de bits para la FSK binaria no coherente, que se presenta en McDonough y Whalen [102], y el correspondiente para el corrimiento de fase por llaveo diferencial que se presenta en Arthurs y Dym [189], incluyen el uso de la distribución de Rician. Ésta surge cuando resulta de interés la envolvente de una onda seno más el ruido gaussiano aditivo; vea el capítulo 1 para una explicación de la distribución de Rician. Las deducciones que se presentan en la sección 6.6 evitan las complicaciones que se enfrentan en el método estándar.

12. El receptor óptimo para el corrimiento de fase por llaveo se explica en Simon y Divsalar [372].

13. Con respecto a una explicación técnica de diversos tipos de módems, subrayando sus características operativas, vea los libros de Lewart [90] y Hold.

14. En un artículo de dos partes de Wei [398], la codificación diferencial se aplica en la codificación de canal convolucional. Ahí se describen varios codificadores convolucionales de ocho estados, que producen códigos que son transparentes a las rotaciones de los elementos de la señal. En particular, en la parte II del artículo, Wei describe reglas y procedimientos de diseño para un código convolucional invariante rotacionalmente en 90 grados que se ha adoptado para utilizarse en el módem V.32 con codificación de enramado.

15. El muestreo no uniforme de señales limitadas en banda se explica en el artículo de Yen [408]. Los resultados principales que se obtienen en ese artículo están contenidos en cuatro teoremas generalizados. La ecuación (6.188) se basa en el teorema III del artículo de Yen.

    En el artículo de Kalet, Mazo y Saltzberg [293], este teorema particular obtenido por Yen se utiliza para formular los criterios fundamentales del diseño del módem digital bidireccional; consulte también la patente de Ayanoglu y otros [193].

16. En lo referente a una explicación de la segunda realización de un módem digital, vea el artículo de Humblet y Troulis [281].

17. En cuanto a una descripción detallada del módem V.34 vea Forney y otros [249]. Los códigos de enramado que se utilizan en el módem V.34 se deben a Wei [398, 399].

18. La idea de la modulación multicanal puede rastrearse hasta los primeros trabajos de Chang [216], Saltzberg [361] y Weinstein y Ebert [401]. En Kalet [292] se presenta un tratamiento matemático de la optimización de la modulación por multitonos para un canal lineal con interferencia de intersímbolo severa. Sin embargo, fue el trabajo de Cioffi y colaboradores lo que condujo a la estandarización de los multitonos discretos (DMT) para líneas digitales asimétricas de suscriptor; en cuanto a los detalles, vea Ruiz y otros [358], Chow y Cioffi, la sección 7.2 del libro de Starr y otros [147], y el capítulo 11 de Cioffi [30]. El problema 6.44 se adaptó de Cioffi [30].

19. El *método de los multiplicadores de Lagrange* para determinar los valores extremos de la función

$$y = f(x)$$

    sujeta a la restricción

$$\varphi(x) = 0$$

    se desprende del siguiente teorema: una condición necesaria y suficiente para un extremo de una función continuamente diferenciable $f(x)$ es que su diferencial con respecto a x se anula en el punto crítico (es decir, el máximo y el mínimo) de la función. Por tanto, en los puntos críticos de $f(x)$ tenemos

$$\frac{\partial f}{\partial x}dx = 0 \tag{1}$$

    Además, puesto que $\varphi(x) = 0$, su diferencial también se anula, según se indica por medio de

$$\frac{\partial \varphi}{\partial x}dx = 0 \tag{2}$$

    Por tanto, al multiplicar (2) por un parámetro λ y luego sumar el resultado a (1) obtenemos

$$\left(\frac{\partial f}{\partial x} + \lambda\frac{\partial \varphi}{\partial x}\right)dx = 0$$

    Puesto que dx es un incremento independiente, deducimos de inmediato

$$\frac{\partial}{\partial x}(f(x) + \lambda\,\varphi(x)) = 0$$

    Esta ecuación es un enunciado matemático de método de los multiplicadores de Lagrange. El parámetro λ recibe el nombre de *multiplicador de Lagrange*. El material que se presenta en esta nota se basó en Sokolnikoff y Redheffer ([144], pp. 341-344).

20. En la transformada discreta de Fourier (DFT), tanto la entrada como la salida consisten en secuencias de números definidos en puntos espaciados de modo uniforme en el tiempo y en la frecuencia, respectivamente. Esta característica hace que la DFT sea idealmente adecuada para el cálculo numérico utilizando el *algoritmo de la transformada rápida de Fourier (FFT)*. Los algoritmos de la FFT son eficientes debido a que usan un número bastante reducido de operaciones aritméticas en comparación con el cómputo de fuerza bruta de la DFT. Básicamente, un algoritmo FFT alcanza su eficiencia de cómputo siguiendo una estrategia de "divide y vencerás", en la cual el cálculo DFT original se descompone sucesivamente en cálculos DFT más pequeños. En el caso de una DFT de

N puntos y $N = 2^L$, el algoritmo FFT requiere $L = \log_2 N$ etapas de cómputo, con cada etapa incluyendo multiplicaciones y adiciones complejas de orden N. En lo que respecta a una discusión detallada del algoritmo FFT consulte Oppenheim y Schafer (capítulo 9).

21. Un panorama de las líneas digitales de alta velocidad de suscriptor (VDSL) se presenta en el artículo de Cioffi y otros [30]; en éste se incluye también una explicación comparativa de las VDSL y de los módems de banda de voz.

22. Para una explicación del OFDM y sus aplicaciones, vea Casas y Leung [214], LeFloch y otros [245] y Zou y Wu [413]. Con respecto a notas instructivas sobre el OFDM y una lista amplia de referencias, vea Cimini y Li [220].

23. En lo relativo a descripciones detalladas de la recuperación de fase y recuperación de la temporización del símbolo utilizando sistemas de sincronización clásicos, vea Stiffler [150], Lindsey [92] y Lindsey y Simon ([94], capítulos 2 y 9).

    Para un tratamiento moderno de los sistemas de sincronización destacando el uso de algoritmos de procesamientos de señales en tiempo discreto, consulte Mengali y D'Andrea [104], Meyr, Moeneclaey y Fechtel [108].

24. La ecuación (6.293) sobre la cota modificada de Cramér-Rao para la recuperación de fase se obtiene en Mengali y D'Andrea [104].

25. Saltzberg [360] explica cómo los desempeños de la CAP y los DMT resultan afectados por los deterioros del canal y las imperfecciones del sistema en el contexto de la aplicación de la ADSL. Los deterioros-imperfecciones que se consideran ahí incluyen el ruido del pulso, la interferencia de banda angosta (por ejemplo el ingreso de RF proveniente de la transmisión por el aire de radio AM), inestabilidad de temporización provocada por sincronización imperfecta y no linealidades del sistema.

PROBLEMAS

Corrimiento de amplitud por llaveo

6.1 En la versión de activado-desactivado por llaveo de un sistema ASK, el símbolo 1 se representa transmitiendo una portadora senoidal de amplitud $\sqrt{2E_b/T_b}$, donde E_b es la energía de la señal por bit y T_b es la duración de bit. El símbolo 0 se representa desactivando la portadora. Suponga que los símbolo 1 y 0 ocurren con igual probabilidad.

En un canal AWGN, determine la probabilidad promedio de error para este sistema ASK bajo los siguientes escenarios:

a) Recepción coherente.

b) Recepción no coherente, operando con un elevado valor de la relación de energía de bit a la densidad espectral de ruido E_b/N_0.

Nota: cuando x es grande, la función de Bessel debe ser modificada de primer tipo de orden cero, por lo que puede aproximarse como sigue (vea el apéndice 3):

$$I_0(x) \simeq \frac{\exp(x)}{\sqrt{2\pi x}}$$

Corrimiento de fase por llaveo

6.2 Una señal PSK se aplica a un correlator alimentado con una referencia de fase que se encuentra dentro de φ radianes de la fase exacta de la portadora. Determine el efecto del error de fase φ sobre la probabilidad promedio de error del sistema.

6.3 Considere una malla de fase encadenada compuesta por un multiplicador, filtro de lazo y oscilador controlado por voltaje. Sea la señal aplicada a la entrada del multiplicador una señal PSK definida por

$$s(t) = A_c \cos\left[2\pi f_c t + k_p m(t)\right]$$

donde k_p es la sensibilidad de fase, y la señal de datos $m(t)$ toma el valor $+1$ para el símbolo binario 1 y -1 para el símbolo 0. La salida del oscilador controlado por voltaje es

$$r(t) = A_c \operatorname{sen}[2\pi f_c t + \theta(t)]$$

a) Evalúe la salida del filtro del lazo, suponiendo que este filtro elimina sólo las componentes moduladas con frecuencia de portadora $2f_c$.

b) Muestre que esta salida es proporcional a la señal de datos $m(t)$ cuando entra a una malla de fase encadenada, esto es, $\theta(t) = 0$.

6.4 La componente de señal de un sistema PSK coherente está definida por

$$s(t) = A_c k \operatorname{sen}(2\pi f_c t) \pm A_c \sqrt{1-k^2} \cos(2\pi f_c t)$$

donde $0 \leq t \leq T_b$, y el signo más corresponde al símbolo 1 y el menos al 0. El primer término representa una componente de la portadora incluida con el propósito de sincronizar el receptor con el transmisor.

a) Dibuje un diagrama del espacio de señales para el esquema que se describe aquí; ¿qué observaciones puede hacer acerca de este diagrama?

b) Demuestre que, en presencia de ruido blanco gaussiano aditivo de media cero y densidad espectral de potencia $N_0/2$, la probabilidad promedio de error es

$$P_e = \frac{1}{2}\operatorname{ferc}\left(\sqrt{\frac{E_b}{N_0}(1-k^2)}\right)$$

donde

$$E_b = \frac{1}{2}A_c^2 T_b$$

c) Suponga que 10% de la potencia de la señal transmitida se asigna a la componente de la portadora. Determine el E_b/N_0 que se requiere para originar una probabilidad de error igual a 10^{-4}.

d) Compare el valor de E_b/N_0 con el que se requiere en un sistema PKS convencional con la misma probabilidad de error.

6.5 a) Dada la secuencia de entrada binaria 1100100010, bosqueje las formas de onda de las componentes en fase y en cuadratura de una onda modulada obtenida utilizando la PQSK basada en el conjunto de señales de la figura 6.6.

b) Bosqueje la propia forma de onda QPSK para la secuencia binaria de entrada que se especifica en la parte *a*.

6.6 Sean P_{eI} y P_{eQ} las probabilidades del error de símbolo para los canales en fase y en cuadratura de un sistema de comunicación digital de banda angosta. Demuestre que la probabilidad promedio del error de símbolo para el sistema completo está dada por

$$P_e = P_{eI} + P_{eQ} - P_{eI}P_{eQ}$$

6.7 La ecuación (6.47) es una fórmula aproximada para la probabilidad promedio del error de símbolo correspondiente a la PSK M-aria coherente. Esta fórmula se obtuvo utilizando la cota de unión a la luz del diagrama del espacio de señales de la figura 6.15*b*. Dado que se transmitió el punto de mensaje m_1, demuestre que la fórmula aproximada de la ecuación (6.47) puede obtenerse directamente de la figura 6.15*b*.

6.8 Determine la densidad espectral de potencia de una señal QPSK de compensación producida por una secuencia binaria aleatoria en la cual los símbolos 1 y 0 (representados por ± 1) son igualmente probables, y donde los símbolos en diferentes tramos de tiempo son estadísticamente independientes y se distribuyen en forma idéntica.

6.9 La modulación de banda lateral residual (VSB), expuesta en el capítulo 2, ofrece otro método de modulación para la transmisión de datos pasobanda.

a) En particular, un sistema de transmisión VSB digital puede considerarse como un sistema unidimensional variable en el tiempo y que opera a una velocidad de $2/T$ dimensiones por segundo, donde T es el periodo del símbolo. Justifique la validez de este enunciado.

b) Demuestre que la VSB digital es en realidad equivalente en desempeño a la QPSK de compensación.

6.10 Se aplica la secuencia de datos binarios 01101000 a una DQPSK desfasada $\pi/4$ que está inicialmente en el estado $(\phi_1 = \sqrt{E}, \phi_2 = 0)$ en la figura 6.11*a*. Utilizando las relaciones entre los dibits de entrada y los corrimientos de la fase portadora que se resumen en la tabla 6.2, establezca los estados de fase ocupados por el modulador en respuesta a la secuencia de datos especificados.

6.11 Del mismo modo que en un modulador QPSK ordinario, la salida de un modulador DQPSK desfasada $\pi/4$ puede expresarse en términos de sus componentes en fase y en cuadratura del siguiente modo:

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\,\text{sen}(2\pi f_c t)$$

Formule la componente en fase $s_I(t)$ y la componente en cuadratura $s_Q(t)$ de la señal DQPSK desfasada $\pi/4$. Por tanto, describa un esquema para la generación de señales DQPSK desfasada $\pi/4$.

6.12 Una propiedad interesante de las señales DQPSK desfasada $\pi/4$ es que pueden demodularse utilizando un discriminador de FM. Demuestre la validez de esta propiedad. El discriminador de FM se presenta en el capítulo 2.

6.13 Sea $\Delta\theta_k$ la fase codificada diferencialmente en la DQPSK desfasada $\pi/4$. Los pares de símbolos (I, Q) generados por este esquema pueden definirse como

$$I_k = I_{k-1}\cos(\Delta\theta_k) - Q_{k-1}\,\text{sen}(\Delta\theta_k)$$
$$Q_k = I_{k-1}\,\text{sen}(\Delta\theta_k) + Q_{k-1}\cos(\Delta\theta_k)$$

donde I_k y Q_k son las componentes en fase y en cuadratura que corresponden al k-ésimo símbolo. Demuestre que este par de relaciones puede expresarse simplemente como

$$I_k = \cos\theta_k$$
$$Q_k = \text{sen}\,\theta_k$$

donde θ_k es el ángulo de fase absoluto del k-ésimo símbolo.

Modulación de amplitud en cuadratura

6.14 La figura 6.53 muestra una constelación de señales QAM 240, que puede considerarse como una forma extendida de la constelación cruzada QAM.

a) Identifique la porción de la figura 6.53 que es una constelación cuadrada QAM.

b) Básese en la parte *a* para identificar la parte de la figura 6.53 que es una constelación cruzada QAM.

c) Por tanto, identifique la parte de la figura 6.53 que es una extensión de la constelación cruzada QAM.

6.15 Determine la reducción del ancho de banda de transmisión y la energía promedio de la señal de la QAM 256 comparada con la QAM 64.

6.16 Se van a comparar dos sistemas de transmisión de datos pasobanda. Uno usa PSK 16 y el otro QAM 16. Ambos se requieren para producir una probabilidad promedio del error de símbolo igual a 10^{-3}. Compare los requerimientos de la relación de señal a ruido de estos dos sistemas.

Modulación de amplitud/fase sin portadora (CAP)

6.17 Los esquemas de la CAP bidimensional y de la QAM M-aria están estrechamente relacionados. Efectúe lo siguiente:

a) Dado un sistema QAM, con un número preestablecido de niveles de amplitud, obtenga el sistema CAP equivalente.

b) Realice lo inverso de la parte *a*.

6.18 Demuestre que la densidad espectral de potencia de una señal CAP con un total de L niveles de amplitud está definida por

$$S(f) = \frac{\sigma_A^2}{T}|P(f)|^2$$

donde $|P(f)|$ es el espectro de magnitud del pulso en fase pasobanda $p(t)$; σ_A^2 es la varianza de los símbolos complejos $A_i = a_i + jb_i$, la cual está definida por

$$\sigma_A^2 = \frac{1}{L}\sum_{i=1}^{L}(a_i^2 + b_i^2)$$

6.19 Se le proporciona un espectro de coseno elevado de banda base $G(f)$ correspondiente a cierto factor de caída α. Describa un procedimiento en el dominio de la frecuencia para evaluar el pulso en fase pasobanda $p(t)$ y el pulso en cuadratura $\hat{p}(t)$ que caracterizan a la señal CAP correspondiente.

Corrimiento de frecuencia por llaveo

6.20 Los vectores de señal $\mathbf{s}_1$ y $\mathbf{s}_2$ se emplean para representar los símbolos binarios 1 y 0, respectivamente, en un sistema FSK binario coherente. El receptor decide en favor del símbolo 1 cuando

$$\mathbf{x}^T\mathbf{s}_1 > \mathbf{x}^T\mathbf{s}_2$$

donde $\mathbf{x}^T\mathbf{s}_i$ es el producto interno del vector de observación $\mathbf{x}$ y el vector de la señal $\mathbf{s}_i$, donde $i = 1, 2$. Demuestre que esta regla de decisión es equivalente a la condición $x_1 > x_2$, donde x_1 y x_2 son los elementos del vector de observación $\mathbf{x}$. Suponga que los vectores de señal $\mathbf{s}_1$ y $\mathbf{s}_2$ tienen igual energía.

6.21 Un sistema FSK transmite datos binarios a razón de 2.5×10^6 bits por segundo. Durante el curso de la transmisión, se añade a la señal ruido blanco gaussiano de media cero y densidad espectral de potencia 10^{-20} W/Hz. A falta de ruido, la amplitud de la onda senoidal recibida para el dígito 1 o 0 es de 1 mV. Determine la probabilidad promedio del error de símbolo de las siguientes configuraciones de sistema:

a) FSK binaria coherente
b) MSK coherente
c) FSK binaria no coherente

6.22 a) En un sistema FSK coherente, las señales $s_1(t)$ y $s_2(t)$ que representan respectivamente los símbolos 1 y 0, se definen como

$$s_1(t), s_2(t) = A_c \cos\left[2\pi\left(f_c \pm \frac{\Delta f}{2}\right)t\right], \quad 0 \le t \le T_b$$

Suponiendo que $f_c > \Delta f$, demuestre que el coeficiente de correlación de las señales $s_1(t)$ y $s_2(t)$ está dado aproximadamente por

$$\rho = \frac{\int_0^{Tb} s_1(t)s_2(t)\,dt}{\int_0^{Tb} s_1^2(t)\,dt} \simeq \text{senc}(2\Delta f T_b)$$

b) ¿Cuál es el valor mínimo del corrimiento de frecuencia Δf para el cual las señales $s_1(t)$ y $s_2(t)$ son ortogonales?
c) ¿Cuál es el valor de Δf que minimiza la probabilidad promedio del error de símbolo?
d) Para el valor de Δf que se obtuvo en la parte c, determine el aumento que se requiere en E_b/N_0 para que este sistema FSK coherente tenga el mismo desempeño ante el ruido que un sistema PSK binario coherente.

6.23 Una señal FSK binaria con *fase discontinua* está definida por

$$s(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T_b}}\cos\left[2\pi\left(f_c + \dfrac{\Delta f}{2}\right)t + \theta_1\right] & \text{para el símbolo 1} \\ \sqrt{\dfrac{2E_b}{T_b}}\cos\left[2\pi\left(f_c - \dfrac{\Delta f}{2}\right)t + \theta_2\right] & \text{para el símbolo 0} \end{cases}$$

donde E_b es la energía de la señal por bit, T_b es la duración de bit y θ_1 y θ_2 son valores de muestreo de variables aleatorias distribuidas uniformemente sobre el intervalo de 0 a 2π. En efecto, los dos osciladores que suministran las frecuencias transmitidas $f_c \pm \Delta f/2$ operan independientemente entre sí. Suponga que $f_c >> \Delta f$.

a) Evalúe la densidad de potencia espectral de la señal FSK.

b) Demuestre que para frecuencias lejanas eliminadas de la frecuencia de la portadora f_c, la densidad espectral de potencia disminuye como el inverso del cuadrado de la frecuencia.

6.24 Elabore un diagrama de bloques para la generación de la señal FSK de Sunde $s(t)$ con fase continua utilizando la representación que se indica en la ecuación (6.104), misma que se repite aquí:

$$s(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(\frac{\pi t}{T_b}\right) \cos(2\pi f_c t) \mp \sqrt{\frac{2E_b}{T_b}} \operatorname{sen}\left(\frac{\pi t}{T_b}\right) \operatorname{sen}(2\pi f_c t)$$

6.25 Explique las similitudes entre la MSK y la QPSK, así como los rasgos que las distinguen.

6.26 Hay dos formas de detectar una señal MSK. Una consiste en usar un receptor coherente para tomar en cuenta completamente el contenido de información de la fase de la señal MSK. Otra forma es usar un receptor no coherente y descartar la información de la fase. El segundo método ofrece la ventaja de la facilidad de la puesta en práctica, a costa de desempeño degradado ante el ruido. ¿En cuántos decibeles tenemos que aumentar la relación de la energía de bit a la densidad de ruido E_b/N_0 en el segundo caso de modo que se produzca una probabilidad promedio del error de símbolo igual a 10^{-5} en ambos casos?

6.27 a) Dibuje las formas de onda de las componentes en fase y en cuadratura de la señal MSK en respuesta a la secuencia binaria de entrada 1100100010.

b) Dibuje la propia forma de onda MSK para la secuencia binaria especificada en la parte *a*.

6.28 Un flujo de datos sin retorno a cero (de niveles de amplitud ± 1) se hace pasar por un filtro pasobajas cuya respuesta al impulso está definida por la función gaussiana

$$h(t) = \frac{\sqrt{\pi}}{\alpha} \exp\left(-\frac{\pi^2 t^2}{\alpha^2}\right)$$

donde α es un parámetro de diseño definido en términos del ancho de banda de 3 dB del filtro mediante

$$\alpha = \sqrt{\frac{\log 2}{2}} \frac{1}{W}$$

a) Demuestre que la función de transferencia del filtro está definida por

$$H(f) = \exp(-\alpha^2 f^2)$$

Por tanto, demuestre que el ancho de banda de 3 dB del filtro es en realidad igual a W. Puede recurrir a la tabla A6.3 sobre los pares de transformada de Fourier.

b) Demuestre que la respuesta del filtro al pulso rectangular de amplitud unitaria y duración T centrado en el origen está definida por la ecuación (6.135).

6.29 Grafique la forma de onda de un modulador GMSK producida en respuesta a la secuencia binaria 1101000, suponiendo el uso de un producto tiempo-ancho de banda $WT_b = 0.3$. Compare sus resultados con los del ejemplo 6.5.

6.30 Resuma las similitudes y diferencias entre la MSK estándar y las señales MSK de filtrado gaussiano.

Receptores no coherentes

6.31 En la sección 6.8 obtuvimos la fórmula para la tasa de error de bits de la FSK binaria no coherente como un caso especial de modulación ortogonal no coherente. En este problema repasamos este

tema. Como antes, suponemos que el símbolo binario 1 representa que se transmite la señal $s_1(t)$. De acuerdo con el material que se presenta en la sección 6.8, señalamos lo siguiente:

- La variable aleatoria L_2 representada por el valor de muestreo l_2 de la ecuación (6.164) exhibe una distribución de Rayleigh.
- La variable aleatoria L_1 representada por el valor de muestreo l_1 de la ecuación (6.170) exhibe una distribución de Rician.

Las distribuciones de Rayleigh y Rician se explican en el capítulo 1. Empleando las distribuciones de probabilidad definidas en ese capítulo, obtenga la fórmula de la ecuación (6.181) para la tasa de error de bits de la FSK binaria no coherente.

6.32 La figura P6.32*a* muestra un receptor no coherente utilizando un filtro de acoplamiento para la detección de una señal senoidal de frecuencia conocida pero fase aleatoria, en la presencia de ruido blanco aditivo gaussiano. Una realización alternativa de este receptor es su mecanización en el dominio de la frecuencia como un *receptor analizador de espectro*, como en la figura P6.32*b*, donde el correlator calcula la función de autocorrelación de tiempo finito $R_x(\tau)$ definida por

$$R_x(\tau) = \int_0^{T-\tau} x(t)x(t+\tau)\,dt, \quad 0 \le \tau \le T$$

Demuestre que la salida del detector de envolvente de ley cuadrática en el tiempo $t = T$ en la figura P6.32*a* es el doble de la salida espectral del transformador de Fourier muestreado a la frecuencia $f = f_c$ en la figura P6.32*b*.

FIGURA P6.32

6.33 La secuencia binaria 1100100010 se aplica al transmisor DPSK de la figura 6.43*a*.

a) Dibuje la forma de onda resultante a la salida del transmisor.

b) Aplicando esta forma de onda al receptor DPSK de la figura 6.43*b*, demuestre que, a falta de ruido, la secuencia binaria original se reconstruye a la salida del receptor.

6.34 La *PSK M-aria diferencial* es la extensión M-aria de la DPSK binaria. El ángulo de fase presente θ_n del modulador en el tiempo n del símbolo se determina recursivamente mediante la relación

$$\theta_n = \theta_{n-1} + \left(\frac{2\pi}{M}\right) m_n, \quad \text{módulo } 2\pi$$

donde θ_{n-1} es el ángulo de fase previo y $m_n \in \{0, 1, \ldots, M-1\}$ es la entrada presente del modulador. La probabilidad del error de símbolo para este esquema de modulación M-ario está dada aproximadamente por

$$P_e \simeq \text{ferc}\left(\sqrt{\frac{2E}{N_0}}\,\text{sen}\left(\frac{\pi}{2M}\right)\right), \quad M \ge 4$$

donde se supone que E/N_0 es grande.

a) Determine el factor mediante el cual la energía transmitida por símbolo tendría que incrementarse en la PSK M-aria diferencial para alcanzar la misma probabilidad del error de símbolo que la PSK M-aria coherente relativa a $M \ge 4$.

b) Para $M = 4$, ¿en cuántos decibeles es más pobre el desempeño de la QPSK diferencial que el de la QPSK coherente?

Comparación de los esquemas de modulación digital utilizando una sola portadora

6.35 Se transmiten datos binarios por un enlace de microondas a una tasa de 10^6 b/s y la densidad espectral de potencia del ruido a la entrada del receptor es 10^{-10} W/Hz. Determine la potencia promedio de la portadora para mantener una probabilidad promedio de error $P_e \leq 10^{-4}$ para *a*) PSK binaria coherente y *b*) DPSK.

6.36 Los valores de E_b/N_0 que se requieren para realizar una probabilidad promedio del error de símbolo $P_e = 10^{-4}$ empleando los sistemas PSK binario coherente y FSK (convencional) coherente son, respectivamente, iguales a 7.2 y 13.5. Empleando la aproximación

$$\text{ferc}(u) \simeq \frac{1}{\sqrt{\pi u}} \exp(-u^2)$$

determine la separación en los valores de E_b/N_0 para $P_e = 10^{-4}$, utilizando

a) PSK binaria coherente y DPSK.

b) PSK binaria coherente y QPSK.

c) FSK (convencional) binaria coherente y FSK binaria no coherente.

d) FSK (convencional) binaria coherente y MSK coherente.

6.37 En la sección 6.10 comparamos los desempeños antes del ruido de la PSK binaria coherente, la FSK binaria coherente, la QPSK, la MSK, la DPSK y la FSK no coherente utilizando la tasa de error de bits como la base de comparación. En este problema tomamos un punto de vista diferente y empleamos la probabilidad promedio del error de símbolo, P_e, para efectuar la comparación. Grafique P_e en función de E_b/N_0 para cada uno de estos esquemas y comente sus resultados.

6.38 El *ancho de banda equivalente de ruido* de la señal pasobanda se define como el valor del ancho de banda que satisface la relación

$$2BS(f_c) = P/2$$

donde $2B$ es el ancho de banda equivalente de ruido centrado alrededor de la frecuencia de banda media f_c, $S(f_c)$ es el valor máximo de la densidad espectral de potencia de la señal en $f = f_c$ y P es la potencia promedio de la señal. Demuestre que los anchos de banda equivalentes de ruido de la PSK binaria, la QPSK y la MSK son como sigue:

Tipo de modulación	*Ancho de banda del ruido/Tasa de bits*
PSK binaria	1.0
QPSK	0.5
MSK	0.62

Nota: puede utilizar las integrales definidas en la tabla A6.10. En el apéndice 2 se presenta una explicación del ancho de banda equivalente del ruido.

Módems de banda de voz

6.39 **a)** Refiérase al codificador diferencial que se usa en la figura 6.48*a*. La tabla 6.10 define los cambios de fase inducidos en el módem V.32 variando los dibits de entrada. Amplíe esta tabla incluyendo los valores correspondientes anterior y presente de la salida del codificador diferencial. Advierta que para cada dibit de entrada $Q_{1,n}Q_{2,n}$, hay cuatro valores posibles para el dibit codificado diferencialmente $I_{1,n}I_{2,n}$, y de igual modo para su valor previo $I_{1,n-1}I_{2,n-1}$.

b) El cuadribit presente aplicado al módem V.32 con codificación no redundante es 0001. La salida previa del módem es 01. Determine la salida de palabra de código que produce el módem y sus coordenadas.

6.40 El estándar de módem V.32 con codificación no redundante utiliza una constelación QAM 16. Las especificaciones del modelo son las siguientes:

Frecuencia de portadora = 1800 Hz
Velocidad del símbolo = 2400 bauds
Velocidad de datos = 9600 b/s

Calcule *a*) la relación señal a ruido promedio y *b*) la probabilidad promedio del error de símbolo para este módem, suponiendo que $E_{pro}/N_0 = 20$ dB.

Códigos de línea multicanal

6.41 Considere las funciones básicas pasobanda que se definen en la ecuación (6.196), donde la propia $\phi(t)$ se define en la ecuación (6.197). Demuestre la validez de las propiedades 1, 2 y 3 de estas funciones básicas pasobanda mencionadas en las páginas 434 y 435.

6.42 La solución con la analogía del llenado de agua para el problema de carga se define por medio de la ecuación (6.213) sujeta a la restricción de la ecuación (6.210). Empleando este par de relaciones, formule un algoritmo recursivo para el cálculo de la asignación de la potencia de transmisión P entre los N subcanales. El algoritmo debe empezar con *a*) una *relación de señal a ruido* inicial total o suma SNR(i) = 0 para la iteración $i = 0$, y *b*) los subcanales clasificados en términos de la asignación de potencia más pequeña a la más grande.

6.43 La respuesta en magnitud al cuadrado de un canal lineal, denotada por $|H(f)|^2$, se muestra en la figura P6.43. Suponga que la brecha $\Gamma = 1$ y la varianza del ruido $\sigma_n^2 = 1$ en todos los subcanales.

a) Obtenga las fórmulas para las potencias óptimas P_1, P_2 y P_3 asignadas a los tres subcanales de bandas de frecuencia $(0, W_1)$, (W_1, W_2) y (W_2, W).

b) Dado que la potencia de transmisión total $P = 10$, $l_1 = 2/3$ y $l_2 = 1/3$, calcule los valores correspondientes de P_1, P_2 y P_3.

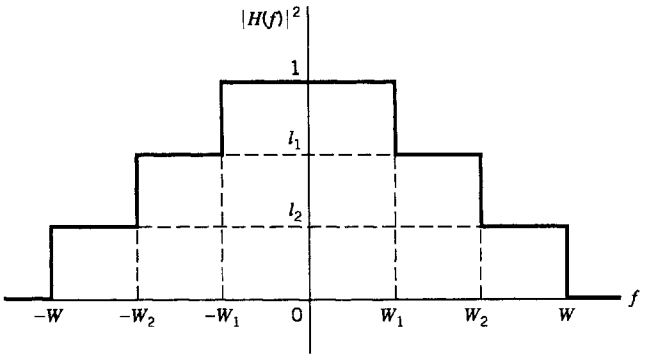


FIGURA P6.43

6.44 En este problema exploramos el uso de la *descomposición de valor singular* (SVD) como una alternativa para la transformada discreta de Fourier para la codificación de vectores. Este método evita la necesidad de un prefijo cíclico, formulándose la matriz del canal como

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_v & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{v-1} & h_v & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & h_0 & h_1 & \cdots & h_v \end{bmatrix}$$

donde la secuencia $h_0, h_1, ..., h_\nu$ denota la respuesta al impulso muestreada del canal. La SDV de la matriz $\mathbf{H}$ está definida por

$$\mathbf{H} = \mathbf{U}[\mathbf{\Lambda} \vdots \mathbf{O}_{N,\nu}]\mathbf{V}^\dagger$$

donde $\mathbf{U}$ es una matriz unitaria de N por N y $\mathbf{V}$ es una matriz unitaria de $(N + \nu)$ por $(N + \nu)$; es decir

$$\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$$
$$\mathbf{V}\mathbf{V}^\dagger = \mathbf{I}$$

donde $\mathbf{I}$ es la matriz identidad y el exponente daga † denota la transposición hermitiana. El símbolo $\mathbf{\Lambda}$ es una matriz diagonal de N por N con valores singulares λ_n, $n = 1, 2, ..., N$. La $\mathbf{O}_{N,\nu}$ es una matriz de ceros de N por ν.

a) Empleando esta descomposición, demuestre que los N subcanales que resultan del uso de la codificación de vectores están descritos matemáticamente por

$$X_n = \lambda_n A_n + W_n$$

X_n es un elemento del producto de matrices $\mathbf{U}^\dagger\mathbf{x}$, donde $\mathbf{x}$ es la señal recibida (salida del canal). La A_n es el símbolo n-ésimo $a_n + jb_n$ y W_n es una variable aleatoria producto del ruido del canal.

b) Demuestre que la relación de señal a ruido para la codificación de vectores según se describe aquí está dada por

$$(\text{SNR})_{\text{codificación vectorial}} = \Gamma\left(\prod_{n=1}^{N^*}\left(1 + \frac{(\text{SNR})_n}{\Gamma}\right)\right)^{1/(N+\nu)} - \Gamma$$

donde N^* es el número de canales en cada uno de los cuales la potencia del transmisor asignada es no negativa, $(\text{SNR})_n$ es la relación señal a ruido del subcanal n y Γ es una brecha preestablecida.

c) Cuando la longitud del bloque N se acerca a infinito, los valores singulares tienden a las magnitudes de la transformada de Fourier del canal. Empleando este resultado, comente acerca de la relación entre la codificación de vectores y los multitonos discretos.

6.45 Compare el desempeño de los DMT y de la CAP con respecto a los siguientes deterioros del canal:

a) Ruido de impulso.

b) Interferencia de banda angosta.

Suponga que 1) los DMT tienen un gran número de subcanales y 2) el sistema CAP no está codificado y su receptor usa un par de filtros adaptables para ponerse en práctica.

6.46 El multiplexado ortogonal por división de frecuencia puede considerarse como una generalización de la FSK M-aria. Valide las razones de este enunciado.

Sincronización

6.47 La figura P6.47 muestra el diagrama de bloques de un *M-ésimo lazo de potencia* en tiempo continuo para la recuperación de fase en un receptor PSK M-ario.

a) Demuestre que la salida de un dispositivo de ley de potencia M-ario contiene un tono de frecuencia Mf_c, donde f_c es la portadora original.

b) El oscilador en la malla de fase encadenada se ajusta a una frecuencia igual a Mf_c. Justifique dicha elección.

c) El lazo de potencia M-ésimo sufre de un problema de ambigüedad de fase en el sentido de que exhibe M ambigüedades de fase en el intervalo $[0, 2\pi]$. Explique cómo puede surgir este problema en el M-ésimo lazo de potencia. ¿Cómo podría usted resolverlo?

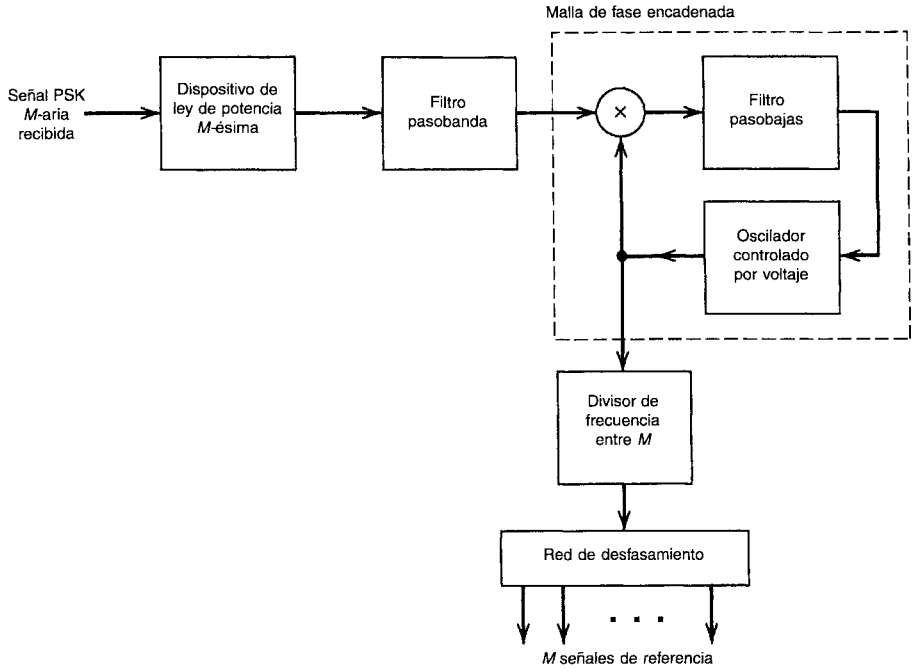


FIGURA P6.47

6.48 **a)** En el algoritmo recursivo de la ecuación (6.272) para la recuperación de fase, la estimación anterior $\hat{\theta}[n]$ y la estimación actualizada $\hat{\theta}[n+1]$ de la fase θ de la portadora se miden en radianes. Analice las unidades en las cuales se miden la señal de error $e[n]$ y el parámetro γ del tamaño del escalón.

b) En el algoritmo recursivo de la ecuación (6.286) para la recuperación de temporización, las señales de control $c[n]$ y $c[n+1]$ son adimensionales. Analice las unidades en las cuales se miden la señal de error $e[n]$ y el parámetro γ del tamaño del escalón.

6.49 Utilizando las definiciones de las ecuaciones (6.264) y (6.265) para $\tilde{x}_k$ y a_k, respectivamente, muestre que el exponente en la función de verosimilitud (a_k, θ, τ) puede expresarse como en la ecuación (6.273).

6.50 En la sección 6.14 estudiamos un esquema sin asistencia de datos para la recuperación de la fase portadora, basado en la función de verosimilitud logarítmica de la ecuación (6.260). En este problema investigamos el uso de esta ecuación para la *recuperación de la fase portadora con asistencia de datos.*

a) Considere un receptor diseñado para un sistema de modulación lineal. Dado que el receptor tiene conocimiento de un preámbulo de la longitud L_0, demuestre que la estimación de máxima verosimilitud de la fase portadora está definida por

$$\hat{\theta} = \arg\left\{\sum_{k=0}^{L_0-1} a_k^* \tilde{x}(k)\right\}$$

donde el preámbulo $\{a_k\}_{k=0}^{L_0-1}$ se conoce como la secuencia de símbolos complejos y $\{\tilde{x}(k)\}_{k=0}^{L_0-1}$ es la envolvente compleja de la señal recibida correspondiente.

b) Utilizando el resultado que se obtuvo en la parte *a*, construya un diagrama de bloques para el estimador de la fase de máxima verosimilitud.

Experimentos de computadora

6.51 El propósito de este experimento de computadora es comparar el efecto de un canal dispersivo sobre las formas de onda generadas por las siguientes técnicas de modulación pasobanda:

a) Corrimiento de fase por llaveo binaria (BPSK)

b) Corrimiento de cuadrifase por llaveo (QPSK)

c) Corrimiento mínimo por llaveo (MSK)

d) MSK gaussiana con producto tiempo-ancho de banda $WT_b = 0.3$

El canal consiste en un *filtro de Butterworth* pasobanda de orden $2N = 10$ y ancho de banda $2B$ de 3 dB centrado en la frecuencia de banda media f_c. El equivalente pasobajas del canal tiene respuesta en magnitud al cuadrado

$$|H(f)|^2 = \frac{1}{1+(f/B)^{2N}}$$

El ancho de banda del canal es variable a fin de ilustrar su efecto sobre la onda modulada filtrada.

Suponiendo el uso de un receptor coherente, grafique las formas de onda de las señales moduladas bajo los siguientes anchos de banda *a*, *b*, *c* y *d* del canal:

i) $2B = 12$ kHz

ii) $2B = 16$ kHz

iii) $2B = 20$ kHz

iv) $2B = 24$ kHz

v) $2B = 30$ kHz

Comente sus resultados.

Sugerencia: para llevar a cabo los cálculos necesarios en este experimento, resulta aconsejable efectuar los cálculos en la banda base realizando la transformación de pasobanda a pasobajas que se describe en el apéndice 2.

CAPÍTULO 7

MODULACIÓN POR ESPECTRO DISPERSO

Este capítulo presenta una técnica de modulación conocida como modulación por espectro disperso, que es radicalmente diferente a las técnicas de modulación que se abordaron en los capítulos anteriores. En dicha técnica, el ancho de banda del canal y la potencia transmitida se sacrifican por el bien de las comunicaciones seguras.

Específicamente, cubrimos los siguientes temas:

- *Secuencias de dispersión en la forma de secuencias de pseudo ruido, sus propiedades y métodos de generación.*
- *La noción básica de la modulación por espectro disperso.*
- *Los dos tipos de modulación por espectro disperso que se usan por lo común: secuencia directa y salto de frecuencia.*

El material que se presenta en este capítulo es básico para las comunicaciones inalámbricas que utilizan acceso múltiple por división de código, lo cual se aborda en el capítulo 8.

7.1 *Introducción*

Un tema de interés principal en el estudio de las comunicaciones digitales, como se consideró en los capítulos 4, 5 y 6, es proveer el uso eficiente de ancho de banda y de potencia. A pesar de la importancia de estos dos recursos de comunicación fundamentales, existen situaciones en las que es necesario sacrificar esta eficiencia para cubrir otros objetivos de diseño. Por ejemplo, es posible que se requiera que el sistema proporcione una forma de comunicación *segura* en ambientes *hostiles*, de tal modo que la señal transmitida no la detecten o reconozcan con facilidad escuchas indeseables. Este requerimiento se obtiene a partir de una clase de técnicas de transmisión de señales conocida colectivamente como *modulación por espectro disperso*.

La ventaja principal del sistema de comunicación de espectro disperso es su capacidad para rechazar la *interferencia*, ya sea si se intenta transmitir simultáneamente por el canal la interferencia *no intencional* de otro usuario o si la interferencia *intencional* de un transmisor hostil intenta perturbar la transmisión.

La definición de la modulación por espectro disperso[1] puede enunciarse en dos partes:

1. El espectro disperso es un medio de transmisión en el cual la secuencia de datos ocupa un ancho de banda en exceso del ancho mínimo necesario para enviarla.
2. La dispersión del espectro se consigue antes de la transmisión mediante el uso de un código que es independiente de la secuencia de datos. El mismo código se utiliza en el

receptor (operando en sincronía con el transmisor) para desdispersar la señal recibida de manera que sea posible recuperar la secuencia original de datos.

Si bien las técnicas de modulación estándar como la modulación de frecuencia y la modulación por codificación de pulsos satisfacen la parte 1 de esta definición, no son técnicas de espectro disperso, debido a que no cumplen con la parte 2 de la misma.

La modulación por espectro disperso se desarrolló originalmente para aplicaciones militares, donde la resistencia a la perturbación (interferencia) es de interés fundamental. Sin embargo, existen aplicaciones civiles que también se benefician de las características únicas de la modulación de espectro disperso. Por ejemplo, puede utilizarse para ofrecer *rechazo de multitrayectoria* en un ambiente de radio móvil terrestre. Incluso otra aplicación corresponde a las comunicaciones de *acceso múltiple* en las cuales se requieren varios usuarios independientes para compartir un canal común sin un mecanismo de sincronización externo; aquí, por ejemplo, es posible mencionar un ambiente de radio terrestre que incluya vehículos móviles que deben comunicarse con una estación central. Esta aplicación se ampliará en el capítulo 8.

En este capítulo explicamos los principios de la modulación por espectro disperso, subrayando las técnicas de secuencia directa y salto de frecuencia. En una técnica de *espectro disperso de secuencia directa* se usan dos etapas de modulación. Primero, la secuencia de datos entrante se utiliza para modular un código de banda ancha. Este código transforma la secuencia de datos de banda angosta en una señal de banda ancha similar al ruido. La señal de banda ancha que se produce se somete a una segunda modulación utilizando una técnica de corrimiento de fase por llaveo. En una técnica de *espectro disperso de salto de frecuencia*, en cambio, el espectro de una portadora modulada por datos se ensancha cambiando la frecuencia de la portadora de una manera pseudoaleatoria. En su operación, ambas técnicas se fundamentan en la disponibilidad de un código de dispersión semejante al ruido conocido como una *secuencia pseudoaleatoria* o *de pseudo ruido*. Puesto que una secuencia de este tipo resulta básica para la operación de la modulación por espectro disperso, es lógico que empecemos nuestro estudio describiendo la generación y las propiedades de las secuencias de pseudo ruido.

7.2 *Secuencias de pseudo ruido*

Una *secuencia de pseudo ruido* es una secuencia binaria periódica con una forma de onda similar al ruido que suele generarse por medio de un *registro de corrimiento realimentado*, cuyo diagrama de bloques general se presenta en la figura 7.1 y que consiste en un *registro de corrimiento* ordinario formado por m flip-flops (etapas de memoria de dos estados) y un *circuito lógico*, mismos que se interconectan en un circuito *realimentado* de lazo múltiple. Los flip-flops en el registro de corrimiento se regulan por medio de un *reloj temporizador*. En cada punto (tic) del reloj, el *estado* de cada flip-flop se cambia al del siguiente en la línea. Con cada pulso del reloj el circuito lógico calcula una función booleana de los estados de los flip-flops. El resultado se realimenta luego

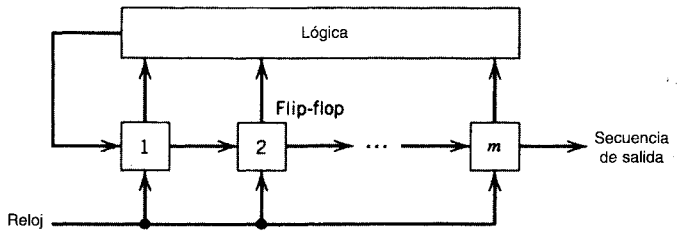


FIGURA 7.1 Registro de corrimiento realimentado.

como la entrada al primero de los flip-flops, evitando de esa manera que el registro de corrimiento se vacíe. La frecuencia de pseudo ruido generada de esa manera se determina por medio de la longitud m del registro de corrimiento, su estado inicial y la lógica de realimentación.

Dejemos que $s_j(k)$ denote el estado del j-ésimo flip-flop después del pulso k-ésimo de reloj; este estado puede representarse mediante el símbolo 0 o 1. El estado del registro de corrimiento después del k-ésimo pulso se define entonces mediante el conjunto $\{s_1(k), s_2(k), \ldots, s_m(k)\}$, donde $k \geq 0$. En el estado inicial, k es cero. De acuerdo con la definición de un registro de corrimiento, tenemos

$$s_j(k+1) = s_{j-1}(k), \quad \begin{cases} k \geq 0 \\ 1 \leq j \leq m \end{cases} \tag{7.1}$$

donde $s_0(k)$ es la entrada aplicada al primer flip-flop después del k-ésimo pulso de reloj. Según la configuración que se describe en la figura 7.1, $s_0(k)$ es una función booleana de los estados individuales $s_1(k), s_2(k), \ldots, s_m(k)$. Para una longitud especificada m, esta función booleana determina en forma única la secuencia de estados subsecuentes y, por tanto, la secuencia de pseudo ruido que se produce a la salida del flip-flop final en el registro de corrimiento. Con el número total m de flip-flops, el número de estados de ruido de este registro de corrimiento es a lo mucho 2^m. Se concluye, en consecuencia, que la frecuencia de pseudo ruido generada por un registro de corrimiento realimentado debe a la larga volverse *periódica* con un periodo de a lo sumo 2^m.

Se dice que un registro de corrimiento será *lineal* cuando la lógica de realimentación consista por completo en *sumadores módulo 2*. En ese caso, *no* se permite el *estado cero* (por ejemplo, el estado en el cual todos los flip-flops se encuentran en el estado cero). Lo señalamos de esa manera debido a que para un estado cero, la entrada $s_0(k)$ producida por la lógica sería 0, el registro de corrimiento permanecería entonces en el estado cero y, por tanto, la salida consistiría enteramente de ceros. De tal modo, el periodo de una secuencia de pseudo ruido producida por un registro de corrimiento realimentado con m flip-flops no puede exceder de $2^m - 1$. Cuando el periodo es exactamente $2^m - 1$, la secuencia de pseudo ruido se denomina *secuencia de longitud máxima* o simplemente *secuencia m*.

▶ Ejemplo 7.1

Considere el registro de corrimiento realimentado lineal que se muestra en la figura 7.2, el cual incluye flip-flops. La entrada s_0 aplicada al primer flip-flop es igual a la suma módulo 2 de s_1 y s_3. Se supone que el estado inicial del registro de corrimiento es 100 (leyendo los contenidos de los tres flip-flops de izquierda a derecha). En ese caso, la sucesión de estados sería como sigue:

100, 110, 111, 011, 101, 010, 001, 100, ...

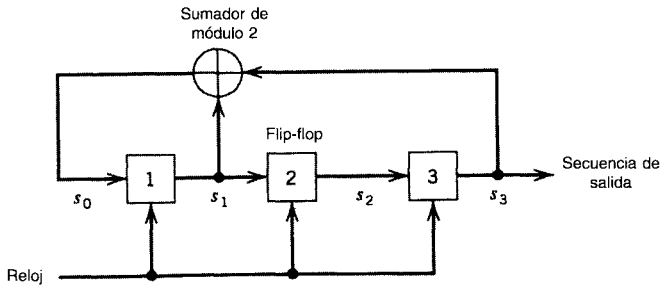


FIGURA 7.2 Generador de secuencia de longitud máxima para $m = 3$.

La secuencia de salida (la última posición de cada estado del registro de corrimiento) es entonces

$$00111010\ldots$$

la cual se repite a sí misma con periodo $2^3 - 1 = 7$.

Note que la elección de 100 como entrada inicial es arbitraria. Cualquiera de los otros seis estados permisibles serviría igualmente bien como un estado inicial. La secuencia de salida que resulta experimentaría simplemente un corrimiento cíclico. ◀

■ PROPIEDADES DE LAS SECUENCIAS DE LONGITUD MÁXIMA[2]

Las secuencias de longitud máxima cuentan con muchas de las propiedades que posee una verdadera *secuencia binaria aleatoria*. Esta última constituye una secuencia en la cual la presencia del número binario 1 o 0 es igualmente probable. Algunas de las propiedades de las secuencias de longitud máxima son las siguientes:

1. En cada periodo de una secuencia de longitud máxima, el número de unos es siempre uno más que el número de ceros. Esto se denomina la *propiedad de balance*.

2. Entre las corridas de unos y ceros en cada periodo de una secuencia de longitud máxima, la mitad de corridas de cada tipo son de longitud 1, un cuarto son de longitud 2, un octavo son de longitud 3 y así sucesivamente, siempre que estas fracciones representen números de corridas significativos. Ésta se conoce como *propiedad de corrida*. Por una "corrida" queremos decir una subsecuencia de símbolos idénticos (unos o ceros) dentro de un periodo de la secuencia. La longitud de esta subsecuencia es la longitud de la secuencia. Para una secuencia de longitud máxima generada por un registro de corrimiento realimentado lineal de longitud m, el número total de corridas es $(N + 1)/2$, donde $N = 2^m - 1$.

3. La función de autocorrelación de una secuencia de longitud máxima es periódica y de valor binario. La anterior se conoce como *propiedad de correlación*.

El periodo de una secuencia de longitud máxima está definido por

$$N = 2^m - 1 \tag{7.2}$$

donde m es la longitud del registro de corrimiento. Considere que los números binarios 0 y 1 de la secuencia se denotan mediante los niveles -1 y $+1$, respectivamente. Sea $c(t)$ la forma de onda resultante de la secuencia de longitud máxima, como se ilustra en la figura 7.3*a* para $N = 7$. El periodo de la forma de onda $c(t)$ es (con base en la terminología que se usa en las secciones subsecuentes)

$$T_b = NT_c \tag{7.3}$$

donde T_c es la duración asignada al símbolo 1 o 0 en la secuencia de longitud máxima. Por definición, la función de autocorrelación de una señal periódica $c(t)$ del periodo T_b es

$$R_c(\tau) = \frac{1}{T_b}\int_{-T_b/2}^{T_b/2} c(t)c(t-\tau)\,dt \tag{7.4}$$

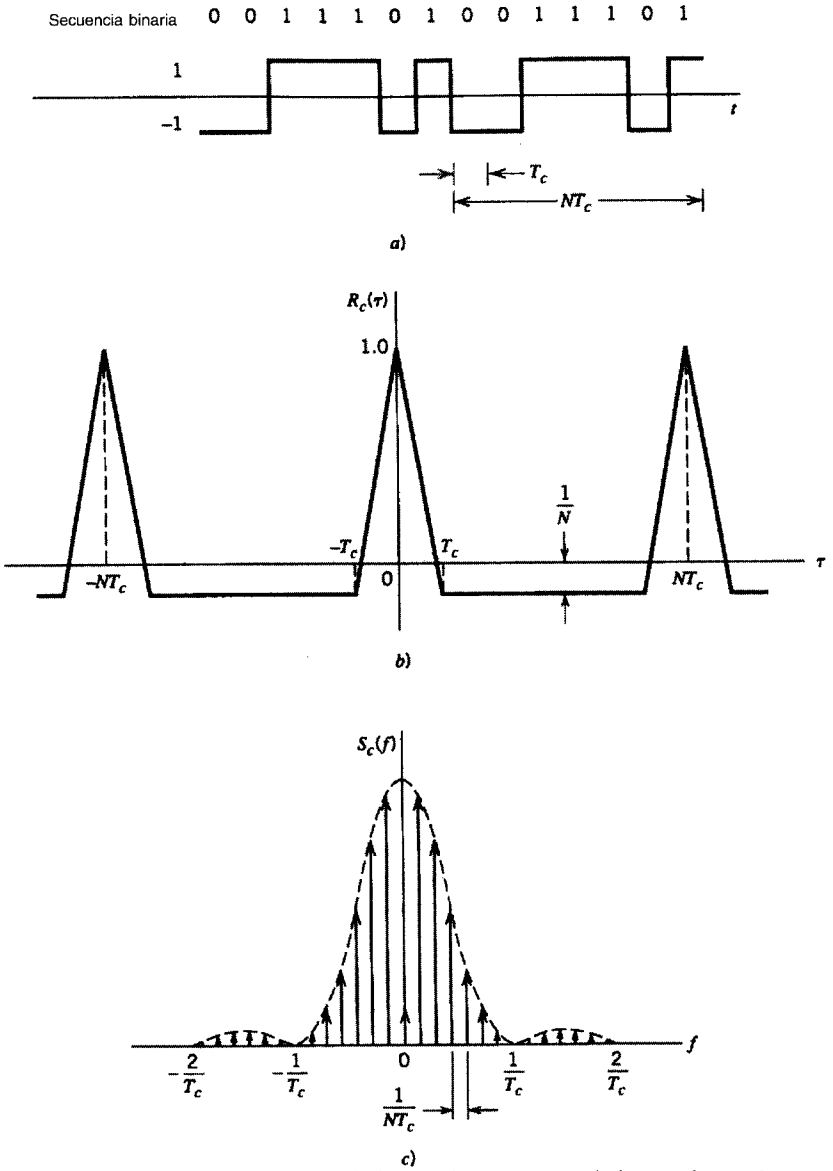


FIGURA 7.3 *a*) Forma de onda de secuencia de longitud máxima para la longitud $m = 3$ o periodo $N = 7$. *b*) Función de autocorrelación. *c*) Densidad espectral de potencia. Las tres partes se refieren a la salida del registro de corrimiento realimentado de la figura 7.2.

donde el retardo τ se encuentra en el intervalo $(-T_b/2, T_b/2)$; la ecuación (7.4) constituye un caso especial de la (1.26). Aplicando esta fórmula a la secuencia de longitud máxima representada por $c(t)$, obtenemos

$$R_c(\tau) = \begin{cases} 1 - \dfrac{N+1}{NT_c}|\tau|, & |\tau| \le T_c \\ -\dfrac{1}{N}, & \text{para el resto del periodo} \end{cases} \tag{7.5}$$

Este resultado se grafica en la figura 7.3*b* para el caso de $m = 3$ o $N = 7$.

De acuerdo con la teoría de la transformada de Fourier sabemos que la periodicidad en el dominio del tiempo se transforma en un muestreo uniforme en el dominio de la frecuencia. Esta interacción entre los dominios del tiempo y la frecuencia se confirma mediante la densidad espectral de potencia de la onda $c(t)$ de longitud máxima. De manera específica, tomando la transformada de Fourier de la ecuación (7.5), obtenemos el espectro de muestreo

$$S_c(f) = \frac{1}{N^2}\delta(f) + \frac{1+N}{N^2}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \text{senc}^2\left(\frac{n}{N}\right)\delta\left(f - \frac{n}{NT_c}\right) \tag{7.6}$$

el cual se grafica en la figura 7.3c para $m = 3$ o $N = 7$.

Al comparar los resultados de la figura 7.3 para una secuencia de longitud máxima con los resultados correspondientes que se muestran en la figura 1.11 para una secuencia binaria aleatoria, podemos hacer las siguientes observaciones:

- En un periodo de secuencia de longitud máxima, la función de autocorrelación $R_c(\tau)$ es algo similar a la de una onda binaria aleatoria.
- Las formas de onda de ambas secuencias tienen la misma envolvente, $\text{senc}^2(fT)$, para sus densidades espectrales de potencia. La diferencia fundamental entre ellas es que en tanto que la secuencia binaria aleatoria tiene una característica de densidad espectral continua, la característica correspondiente de una secuencia de longitud máxima consiste en funciones delta espaciadas por una separación $1/NT_c$ Hz.

Cuando aumenta la longitud m del registro de corrimiento, o equivalentemente, el periodo N de la secuencia de longitud máxima, esta última se vuelve cada vez más similar a la secuencia binaria aleatoria. En realidad, en el límite, las dos secuencias se vuelven idénticas cuando N se hace infinitamente grande. Sin embargo, el precio que se paga por hacer N grande es un aumento en las necesidades de almacenamiento, lo cual impone un límite práctico relativo a cuán grande puede realmente hacerse N.

■ ELECCIÓN DE UNA SECUENCIA DE LONGITUD MÁXIMA

Ahora que entendemos las propiedades de la secuencia de longitud máxima y el hecho de que podemos generarla utilizando un registro de corrimiento realimentado lineal, la pregunta clave que necesitamos plantear es: ¿cómo determinamos la lógica de realimentación en un periodo

TABLA 7.1 ***Secuencias de longitud máxima de longitudes 2-8 del registro de corrimiento***

Longitud del registro de corrimiento, m	*Derivaciones de realimentación*
2*	[2, 1]
3*	[3, 1]
4	[4, 1]
5*	[5, 2], [5, 4, 3, 2], [5, 4, 2, 1]
6	[6, 1], [6, 5, 2, 1], [6, 5, 3, 2]
7*	[7, 1], [7, 3], [7, 3, 2, 1], [7, 4, 3, 2], [7, 6, 4, 2], [7, 6, 3, 1], [7, 6, 5, 2], [7, 6, 5, 4, 2, 1], [7, 5, 4, 3, 2, 1],
8	[8, 4, 3, 2], [8, 6, 5, 3], [8, 6, 5, 2], [8, 5, 3, 1], [8, 6, 5, 1], [8, 7, 6, 1], [8, 7, 6, 5, 2, 1], [8, 6, 4, 3, 2, 1]

deseado N? La respuesta a esta pregunta se encuentra en la teoría de los códigos de control de errores, la cual se aborda en el capítulo 10. La tarea de encontrar la lógica de realimentación requerida se nos facilita particularmente en virtud de las amplias tablas de las fracciones de realimentación necesarias para diversas longitudes de registros de corrimiento que se han compilado en la literatura. En la tabla 7.1 presentamos los conjuntos de derivaciones máximas (realimentación) correspondientes a longitudes $m = 2, 3, \ldots, 8$ de registros de corrimiento.[3] Advierta que cuando aumenta m crece el número de esquemas (códigos) alternativos. Además, para todo conjunto de conexiones de realimentación que se muestran en esta tabla, existe un conjunto "imagen" que genera un código idéntico de longitud máxima, invertido en la secuencia de tiempo.

Los conjuntos particulares que se identifican con un asterisco en la tabla 7.1 corresponden a *secuencias de Mersenne de longitud prima*, para las cuales el periodo N es un número primo.

Ejemplo 7.2

Considere una secuencia de longitud máxima que requiere el uso de un registro de corrimiento lineal realimentado de longitud $m = 5$. Para las derivaciones de realimentación elegimos el conjunto [5, 2] de la tabla 7.1. La configuración correspondiente del generador de códigos se muestra en la figura 7.4*a*. Suponiendo que el estado inicial es 10000, la evolución de un periodo de la secuencia de longitud máxima generada por este esquema se muestra en la tabla 7.2*a*, donde vemos que el generador vuelve al 10000 inicial después de 31 iteraciones, esto es, el periodo es 31, lo cual concuerda con el valor que se obtuvo a partir de la ecuación (7.2).

Suponga a continuación que elegimos otro conjunto de derivaciones de realimentación de la tabla 7.1, a saber, [5, 4, 2, 1]. De ese modo, el generador de código correspondiente es como se indica en la figura 7.4*b*. Para el estado inicial 10000, encontramos en este caso que la evolución de la secuencia de longitud máxima es de la forma que se ilustra en la tabla 7.2*b*. Aquí de nuevo el generador vuelve al estado inicial 10000 después de 31 iteraciones, y así debe ser. Sin embargo, la secuencia de longitud máxima generada es diferente de la que se muestra en la tabla 7.2*a*.

Naturalmente, el generador de código de la figura 7.4*a* tiene una ventaja sobre el de la figura 7.4*b*, ya que requiere un menor número de conexiones de realimentación. ◀

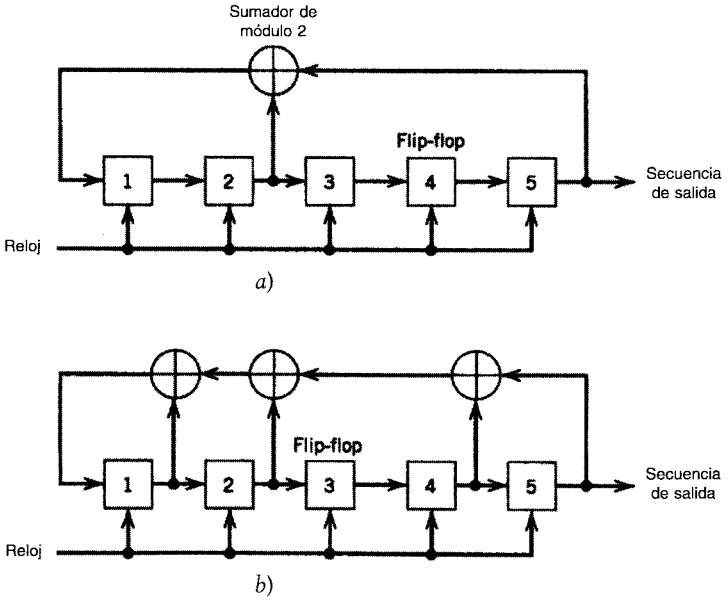


FIGURA 7.4 Dos configuraciones diferentes del registro de corrimiento realimentado de longitud $m = 5$. *a*) Conexiones de realimentación [5, 2]. *b*) Conexiones de realimentación [5, 4, 2, 1].

TABLA 7.2a ***Evolución de la secuencia de longitud máxima generada por el registro de corrimiento realimentado de la figura 7.4a***

Símbolo de realimentación	*Estado del registro de corrimiento*					*Símbolo de salida*
	1	0	0	0	0	
0	0	1	0	0	0	0
1	1	0	1	0	0	0
0	0	1	0	1	0	0
1	1	0	1	0	1	0
1	1	1	0	1	0	1
1	1	1	1	0	1	0
0	0	1	1	1	0	1
1	1	0	1	1	1	0
1	1	1	0	1	1	1
0	0	1	1	0	1	1
0	0	0	1	1	0	1
0	0	0	0	1	1	0
1	1	0	0	0	1	1
1	1	1	0	0	0	1
1	1	1	1	0	0	0
1	1	1	1	1	0	0
1	1	1	1	1	1	0
0	0	1	1	1	1	1
0	0	0	1	1	1	1
1	1	0	0	1	1	1
1	1	1	0	0	1	1
0	0	1	1	0	0	1
1	1	0	1	1	0	0
0	0	1	0	1	1	0
0	0	0	1	0	1	1
1	1	0	0	1	0	1
0	0	1	0	0	1	0
0	0	0	1	0	0	1
0	0	0	0	1	0	0
0	0	0	0	0	1	0
1	1	0	0	0	0	1

Código: **0000101011101100011111001101001**

TABLA 7.2b *Evolución de la secuencia de longitud máxima generada por el registro de corrimiento realimentado de la figura 7.4b*

Símbolo de realimentación	*Estado del registro de corrimiento*					*Símbolo de salida*
	1	0	0	0	0	
1	1	1	0	0	0	0
0	0	1	1	0	0	0
1	1	0	1	1	0	0
0	0	1	0	1	1	0
1	1	0	1	0	1	1
0	0	1	0	1	0	1
0	0	0	1	0	1	0
1	1	0	0	1	0	1
0	0	1	0	0	1	0
0	0	0	1	0	0	1
0	0	0	0	1	0	0
1	1	0	0	0	1	0
0	0	1	0	0	0	1
1	1	0	1	0	0	0
1	1	1	0	1	0	0
1	1	1	1	0	1	0
1	1	1	1	1	0	1
1	1	1	1	1	1	0
0	0	1	1	1	1	1
1	1	0	1	1	1	1
1	1	1	0	1	1	1
0	0	1	1	0	1	1
0	0	0	1	1	0	1
1	1	0	0	1	1	0
1	1	1	0	0	1	1
1	1	1	1	0	0	1
0	0	1	1	1	0	0
0	0	0	1	1	1	0
0	0	0	0	1	1	1
0	0	0	0	0	1	1
1	1	0	0	0	0	1

Código: 0000110101001000101111101100111

7.3 *Una noción del espectro disperso*

Un atributo importante de la modulación por espectro disperso es que ésta puede ofrecer protección contra señales de interferencia (perturbadores) con potencia finita generadas externamente. La señal perturbadora puede consistir en ruido de banda ancha bastante potente o en una forma de onda multitonos que se dirige al receptor con el fin de interrumpir las comunicaciones. La protección contra formas de onda perturbadoras se proporciona al hacer intencionalmente que la señal que contiene la información ocupe un ancho de banda bastante mayor que el mínimo necesario para transmitirla. Esto tiene el efecto de provocar que la señal transmitida asuma una apariencia similar al ruido, de manera que se mezcle con el ruido de fondo. De esa manera se permite que la señal transmitida se propague por el canal sin que la detecte alguien que quizá se encuentre escuchando. Por tanto, es posible pensar en el espectro disperso como un método de "camuflaje" de la señal que contiene la información.

Un método para aumentar el ancho de banda de la secuencia (datos) que contiene la información implica el uso de *modulación*. Considere que $\{b_k\}$ denota una secuencia binaria y $\{c_k\}$ una secuencia de pseudo ruido. Sean las formas de onda $b(t)$ y $c(t)$ sus respectivas representaciones polares sin retorno a cero en términos de dos niveles de igual amplitud y polaridad opuesta, esto es, ± 1. Nos referiremos a $b(t)$ como la señal que contiene la información (datos) y a $c(t)$ como la señal de pseudo ruido. La modulación deseada se logra aplicando la señal de datos de $b(t)$ y la señal de pseudo ruido $c(t)$ en un modulador de producto o multiplicador, como en la figura 7.5*a*. Sabemos de la teoría de la transformada de Fourier que la multiplicación de dos señales produce una señal cuyo espectro es igual a la convolución de los espectros de las dos señales componentes. En consecuencia, si la señal del mensaje $b(t)$ es de banda angosta y la señal de pseudo ruido $c(t)$ es de banda ancha, *la señal producto (modulada) $m(t)$ tendrá un espectro que es casi el mismo que la señal de pseudo ruido de banda ancha.* En otras palabras, en el contexto de nuestra aplicación presente, la secuencia de pseudo ruido desempeña el papel de un *código de dispersión*.

Al multiplicar la señal $b(t)$ que contiene información por la señal de pseudo ruido $c(t)$, cada bit de información se "recorta" en varios incrementos de tiempo pequeños, como se ilustra en las ondas de la figura 7.6. Estos pequeños incrementos de tiempo suelen conocerse como *divisiones*.

En la *transmisión de banda base*, la señal del producto $m(t)$ representa a la *señal transmitida*. Es posible entonces expresar la señal transmitida como

$$m(t) = c(t)b(t) \tag{7.7}$$

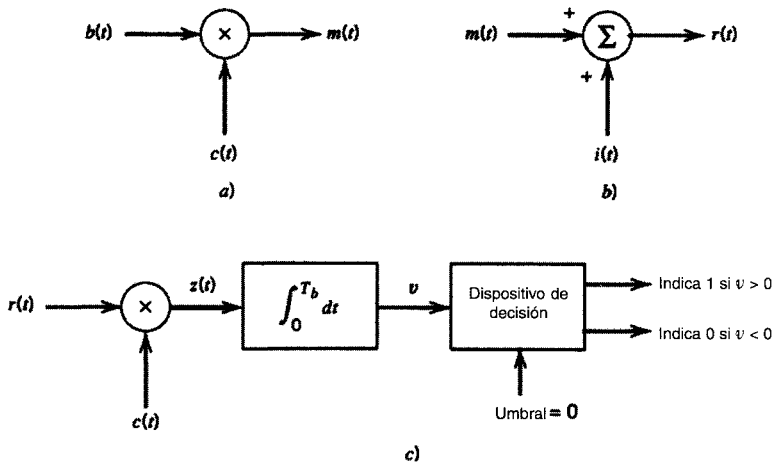


FIGURA 7.5 Modelo idealizado del sistema de espectro disperso de banda base. *a*) Transmisor. *b*) Canal. *c*) Receptor.

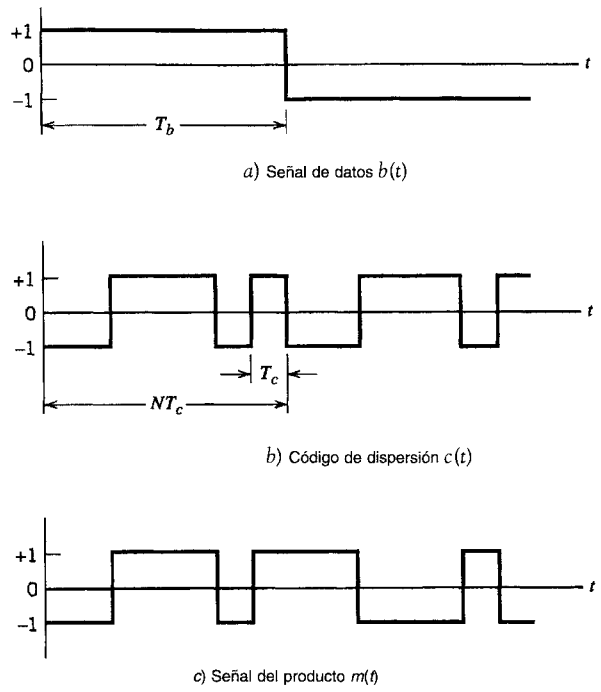


FIGURA 7.6 Ilustración de las formas de onda en el transmisor de la figura 7.5*a*.

La señal recibida $r(t)$ consiste en la señal transmitida $m(t)$ más una *interferencia* aditiva denotada por $i(t)$, como se indica en el modo de canal de la figura 7.5*b*. Por tanto,

$$\begin{aligned} r(t) &= m(t) + i(t) \\ &= c(t)b(t) + i(t) \end{aligned} \tag{7.8}$$

Para recuperar la señal de mensaje original $b(t)$, la señal recibida $r(t)$ se aplica a un *demodulador* que está compuesto por un multiplicador seguido de un integrador y un dispositivo de decisión, como en la figura 7.5*c*. El multiplicador se alimenta con una secuencia de pseudo ruido generada localmente que es una *réplica* exacta del que se usa en el transmisor. Además, suponemos que el receptor opera en *sincronismo* perfecto con el transmisor, lo que significa que la secuencia de pseudo ruido en el receptor está alineada exactamente con la correspondiente al transmisor. La salida del multiplicador en el receptor está dada entonces por

$$\begin{aligned} z(t) &= c(t)r(t) \\ &= c^2(t)b(t) + c(t)i(t) \end{aligned} \tag{7.9}$$

La ecuación (7.9) muestra que la señal de datos $b(t)$ se multiplica dos veces por la señal de ruido $c(t)$, en tanto que la señal indeseable $i(t)$ se multiplica solamente *una vez*. La señal de pseudo ruido $c(t)$ se alterna entre los niveles -1 y $+1$, y la alternancia se destruye cuando ésta se eleva al cuadrado; por consiguiente,

$$c^2(t) = 1 \quad \text{para todo } t \tag{7.10}$$

Por tanto, es posible simplificar la ecuación (7.9) como

$$z(t) = b(t) + c(t)i(t) \tag{7.11}$$

De ese modo observamos, de acuerdo con la ecuación (7.11), que la señal de datos $b(t)$ se reproduce a la salida del multiplicador en el receptor, excepto por el efecto de la interferencia representada por el término aditivo $c(t)i(t)$. La multiplicación de la interferencia $i(t)$ por la señal de pseudo ruido $c(t)$ generada localmente equivale a que el código de dispersión afectará a la interferencia exactamente como lo hizo la señal original en el transmisor. Después de esto observamos que la componente de datos $b(t)$ es de banda angosta, en tanto que la componente espuria $c(t)i(t)$ es de banda ancha. Por ello, al aplicar la salida del multiplicador al filtro de banda base (pasobajas) con un ancho de banda apenas lo suficientemente grande para adaptarse a la señal de recuperación de datos $b(t)$, se filtra la mayor parte de la potencia en la componente espuria $c(t)i(t)$. El efecto de la interferencia $i(t)$ se reduce significativamente en ese caso a la salida del receptor.

En el receptor que se presenta en la figura 7.5c, el efecto del filtrado pasobajas es en realidad ejecutado por el integrador que evalúa el área bajo la señal producida a la salida del multiplicador. La integración se efectúa para el intervalo de bit $0 \leq t \leq T_b$, dado el valor de muestreo v. Por último, el receptor toma una decisión: si v es mayor que el umbral de cero, el receptor indica que se envió el símbolo binario 1 de la secuencia de datos original en el intervalo de $0 \leq t \leq T_b$, y si v es menor que cero, indica que se envió el símbolo 0; si v es exactamente cero, el receptor supone al azar a favor de 1 o 0.

En resumen, el uso de un código de dispersión (con propiedades pseudoaleatorias) en el transmisor produce una señal transmitida de banda ancha que se *asemeja al ruido* en un receptor que *no* tiene conocimiento del código de dispersión. Con base en la explicación que se presenta en la sección 7.2, recordamos que (para una velocidad de datos preestablecida) cuanto más grande hacemos el periodo del código de dispersión, tanto más cercana será la señal transmitida a una onda binaria verdaderamente aleatoria, y tanto más difícil resultará detectarla. Naturalmente, el precio que tenemos que pagar por la protección mejorada contra la interferencia es el aumento en el ancho de banda de transmisión, la complejidad del sistema y el retardo del procesamiento. Sin embargo, cuando nuestro interés principal es la seguridad de la transmisión, el pago de dichos costos no es irrazonable.

7.4 *Espectro disperso de secuencia directa con corrimiento de fase por llaveo binaria coherente*

La técnica del espectro disperso que se describió en la sección anterior se conoce como *espectro disperso de secuencia directa*. La explicación que se presentó se ubica en el contexto de la transmisión de banda base. Para sustentar el uso de esta técnica en la transmisión pasobanda por un canal de satélite, por ejemplo, podemos incorporar el corrimiento de fase por llaveo (PSK) binaria coherente en el transmisor y el receptor, como se muestra en la figura 7.7. El transmisor de la figura 7.7*a* convierte primero la secuencia de datos entrante $\{b_k\}$ en una forma de onda polar sin retorno a cero $b(t)$, a la cual siguen dos etapas de modulación. La primera etapa consta de un modulador de producto o multiplicador con la señal de datos $b(t)$ (representando una secuencia de datos) y la señal de pseudo ruido $c(t)$ (representando la secuencia de pseudo ruido) como entradas. La segunda etapa está integrada por un modulador PSK binario. La señal transmitida $x(t)$ es de ese modo una *señal dispersa de secuencia directa de corrimiento de fase por llaveo binaria* (DS/BPSK). La modulación de fase $\theta(t)$ de $x(t)$ tiene uno de dos valores 0 y π, dependiendo de las polaridades de la señal del mensaje $b(t)$ y de la señal de pseudo ruido $c(t)$ en el tiempo t, de acuerdo con la tabla de verdad de la tabla 7.3.

La figura 7.8 ilustra las formas de onda para la segunda etapa de modulación. Parte de la forma de onda modulada que se muestra en la figura 7.6c se reproduce en la figura 7.8*a*; la forma de onda que se muestra aquí corresponde a un periodo de la secuencia de pseudo ruido. La figura 7.8*b* presenta la forma de onda de una portadora senoidal, y la figura 7.8c muestra la forma de onda DS/BPSK que resulta de la segunda etapa de modulación.

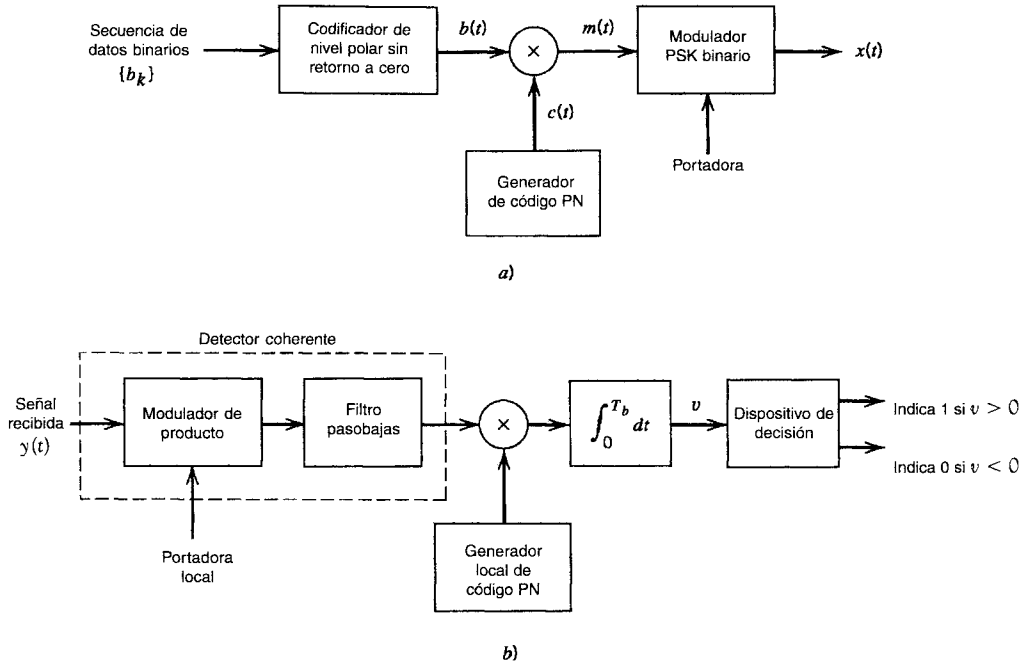


FIGURA 7.7 Corrimiento de fase por llaveo coherente disperso de secuencia directa. *a*) Transmisor. *b*) Receptor.

El receptor, que se muestra en la figura 7.7*b*, está compuesto por dos etapas de demodulación. En la primera, la señal recibida $y(t)$ y una portadora generada localmente se aplican a un modulador de producto seguido por un filtro pasobajas cuyo ancho de banda es igual al de la señal de mensaje original $m(t)$. Esta etapa del proceso de demodulación invierte el corrimiento de fase por llaveo aplicado a la señal transmitida. La segunda etapa de demodulación efectúa la desdispersión del espectro al multiplicar la salida del filtro pasobajas por una réplica generada en forma local de la señal de pseudo ruido $c(t)$, a la que sigue la integración sobre un intervalo de bit $0 \leq t \leq T_b$, y finalmente por una toma de decisión de la manera que se describe en la sección 7.3.

■ MODELO PARA ANÁLISIS

En la forma normal del transmisor, ilustrado en la figura 7.7*a*, la dispersión del espectro se lleva a cabo antes de la modulación de fase. Con fines de análisis, encontramos más conveniente intercambiar el orden de estas operaciones, como se indica en el modelo de la figura 7.9. Se nos

TABLA 7.3 *Tabla de verdad para la modulación de fase* $\theta(t)$, *radianes*

		Polaridad de la secuencia de datos b(t) *al tiempo* t	
		+	−
Polaridad de la secuencia PN c(t) *al tiempo* t	+	0	π
	−	π	0

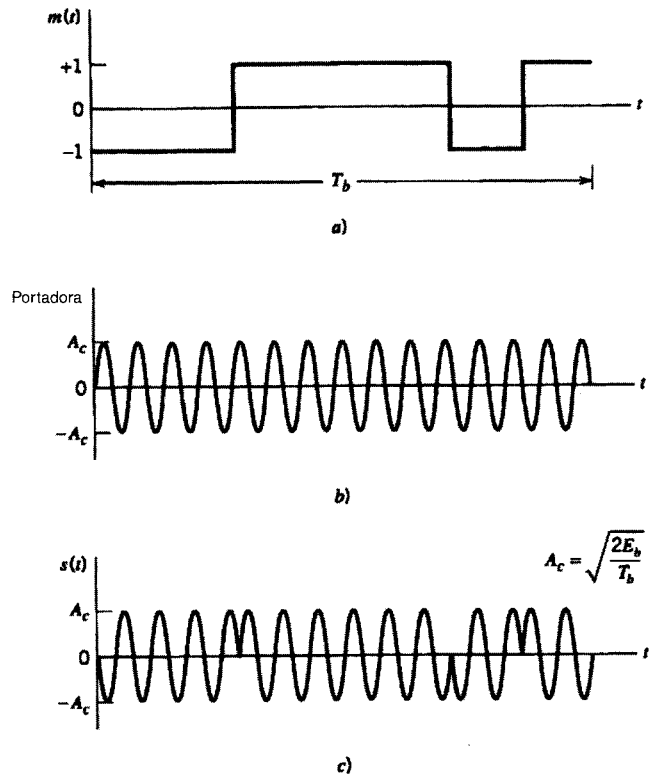


FIGURA 7.8 *a*) Señal del producto $m(t) = c(t)b(t)$. *b*) Portadora senoidal. *c*) Señal DS/BPSK.

permite hacer esto debido a que la dispersión del espectro y el corrimiento de fase por llaveo binaria son ambas operaciones lineales; lo mismo sucede con respecto a la demodulación de fase y a la desdispersión del espectro. No obstante, para que sea factible el intercambio de operaciones, es importante sincronizar la secuencia de datos entrantes y la secuencia de pseudo ruido. El modelo de la figura 7.9 incluye también representaciones del canal y el receptor. En este modelo, se supone que la interferencia $j(t)$ limita el desempeño, de modo que puede ignorarse el efecto del ruido del canal. Por tanto, la salida del canal está dada por

$$\begin{aligned} y(t) &= x(t) + j(t) \\ &= c(t)s(t) + j(t) \end{aligned} \tag{7.12}$$

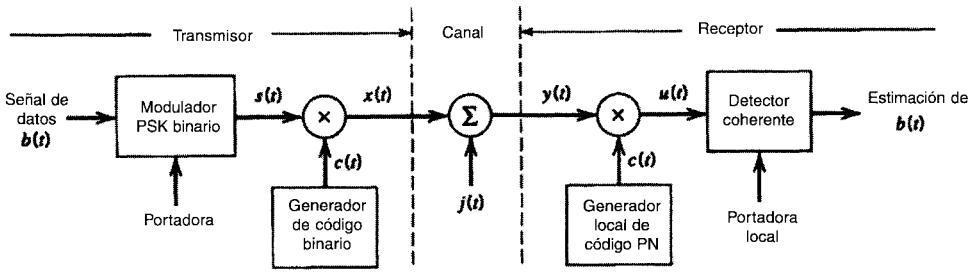


FIGURA 7.9 Modelo del sistema disperso PSK binario de secuencia directa.

donde $s(t)$ es la señal PSK binaria y $c(t)$ es la señal de pseudo ruido. En el modelo del canal que se incluye en la figura 7.9, la señal de interferencia se denota mediante $j(t)$. Esta notación se eligió a propósito con el fin de diferenciarla de la que se utilizó para la interferencia en la figura 7.5*b*. El modelo del canal en la figura 7.9 es pasobanda en contenido espectral, en tanto que el de la figura 7.5*b* está en la forma de la banda base.

En el receptor, la señal recibida $y(t)$ es la primera multiplicada por la señal de pseudo ruido $c(t)$ y produce una salida que es igual a la entrada del detector coherente $u(t)$. De ese modo,

$$\begin{aligned} u(t) &= c(t)y(t) \\ &= c^2(t)s(t) + c(t)j(t) \\ &= s(t) + c(t)j(t) \end{aligned} \tag{7.13}$$

En la última línea de la ecuación (7.13) remarcamos que, por diseño, la señal de pseudo ruido $c(t)$ satisface la propiedad descrita en la ecuación (7.10):

$$c^2(t) = t \quad \text{para todo } t$$

La ecuación (7.13) demuestra que la entrada del detector coherente $u(t)$ consiste en una señal PSK binaria $s(t)$ incorporada en la interferencia aditiva de código modulado que se denota mediante $c(t)j(t)$. La naturaleza modulada de la última componente obliga a que la señal de interferencia (perturbadora) distribuya su espectro de tal manera que la detección de los bits de información a la salida del receptor se ofrezca con una confiabilidad mejorada.

■ Sincronización

Para su operación apropiada, un sistema de comunicación de espectro disperso requiere que la secuencia de pseudo ruido generada localmente utilizada en el receptor para desdispersar la señal recibida esté *sincronizada* con la secuencia de pseudo ruido que se usa para dispersar la señal transmitida en el transmisor.[4] Una solución al problema de sincronización consta de dos partes: *adquisición* y *rastreo*. En la adquisición, o sincronización *gruesa*, los dos códigos de pseudo ruido se alinean hasta dentro de una fracción de la división en un tiempo lo más corto posible. Una vez que se ha adquirido el código de pseudo ruido entrante, ocurre el rastreo o sincronización *fina*. Por lo común la adquisición del pseudo ruido procede en dos etapas: primero, la señal recibida se multiplica por un código de pseudo ruido generado localmente para producir una medida de la *correlación* entre ella y el código de pseudo ruido que se usó en el transmisor. Luego se usa una *regla de decisión* y *estrategia de búsqueda* apropiada para procesar la medida de correlación así obtenida y determinar si los dos códigos están en sincronismo y qué hacer si ese no es el caso. En lo que respecta al rastreo, éste se lleva a cabo utilizando técnicas de cierre de fase similares a las que se utilizaron para la generación local de referencias de portadora coherente. La diferencia principal entre ellas radica en la forma en que se pone en práctica la discriminación de fase.

7.5 *Dimensionalidad del espacio de señales y ganancia del procesamiento*

Habiendo desarrollado una comprensión conceptual de la modulación por espectro disperso y un método para su puesta en práctica, estamos listos para emprender un análisis matemático detallado de la técnica. El método que tenemos en mente se basa en las ideas teóricas del espacio de señales del capítulo 5. En particular, formulamos las representaciones del espacio de señales de la señal transmitida y de la señal de interferencia (perturbadora).

En este contexto consideramos el conjunto de funciones básicas ortogonales:

$$\phi_k(t) = \begin{cases} \sqrt{\frac{2}{T_c}} \cos(2\pi f_c t), & kT_c \le (k+1)T \\ 0, & \text{en otro caso} \end{cases} \tag{7.14}$$

$$\tilde{\phi}_k(t) = \begin{cases} \sqrt{\frac{2}{T_c}} \operatorname{sen}(2\pi f_c t), & kT_c \le (k+1)T \\ 0, & \text{en otro caso} \end{cases} \tag{7.15}$$

$$k = 0, 1, \ldots N-1$$

donde T_c es la *duración de la división* y N es el número de divisiones por bit. Por tanto, es posible describir la señal transmitida $x(t)$ para el intervalo de un bit de información de la manera siguiente:

$$\begin{aligned} x(t) &= c(t)s(t) \\ &= \pm\sqrt{\frac{2E_b}{T_b}} c(t) \cos(2\pi f_c t) \\ &= \pm\sqrt{\frac{E_b}{N}} \sum_{k=0}^{N-1} c_k \phi_k(t), \qquad 0 \le t \le T_b \end{aligned} \tag{7.16}$$

donde E_b es la energía de la señal por bit; el signo más corresponde al bit de información 1, y el signo menos al bit de información 0. La secuencia de código $\{c_0, c_1, \ldots, c_{N-1}\}$ denota la secuencia de pseudo ruido con $c_k = \pm 1$. Por tanto, la señal transmitida $x(t)$ es N-dimensional porque requiere para su representación de un mínimo de N funciones ortogonales.

Considere a continuación la representación de la señal de interferencia (perturbadora) $j(t)$. Idealmente, la perturbación gusta de colocar toda su energía disponible en exactamente el mismo espacio N-dimensional de la señal que la señal transmitida $x(t)$; en otro caso, parte de su energía se desperdicia. Sin embargo, lo máximo que el perturbador puede esperar conocer es el ancho de banda de la señal transmitida. Además, no hay manera de que el perturbador pueda conocer la fase de la señal. Por ello, podemos representar al perturbador por la forma general

$$j(t) \sum_{k=0}^{N-1} j_k \phi_k(t) + \sum_{k=0}^{N-1} \tilde{j}_k \tilde{\phi}_k(t), \qquad 0 \le t \le T_b \tag{7.17}$$

donde

$$j_k = \int_0^{T_b} j(t)\phi_k(t)\, dt, \quad k = 0, 1 \ldots, N-1 \tag{7.18}$$

y

$$\tilde{j}_k = \int_0^{T_b} j(t)\tilde{\phi}_k(t)\, dt, \quad k = 0, 1 \ldots, N-1 \tag{7.19}$$

De ese modo, la interferencia $j(t)$ es $2N$-dimensional; es decir, tiene el doble del número de dimensiones requeridas para representar la señal DS/BPSK $x(t)$ transmitida. En términos de la

representación dada en la ecuación (7.17), es posible expresar la potencia promedio de la interferencia $j(t)$ del modo siguiente:

$$\begin{aligned} J &= \frac{1}{T_b}\int_0^{T_b} j^2(t)\,dt \\ &= \frac{1}{T_b}\sum_{k=0}^{N-1} j_k^2 + \frac{1}{T_b}\sum_{k=0}^{N-1} \tilde{j}_k^2 \end{aligned} \tag{7.20}$$

Además, en virtud de la falta de conocimiento de la fase de la señal, la mejor estrategia que un perturbador puede aplicar es poner igual energía en las coordenadas coseno y seno definidas en las ecuaciones (7.18) y (7.19); por tanto, podemos asumir con seguridad que

$$\sum_{k=0}^{N-1} j_k^2 = \sum_{k=0}^{N-1} \tilde{j}_k^2 \tag{7.21}$$

De modo correspondiente, la ecuación (7.20) se simplifica como

$$J = \frac{2}{T_b}\sum_{k=0}^{N-1} j_k^2 \tag{7.22}$$

Nuestro objetivo es vincular estos resultados determinando las relaciones señal a ruido medidas a la entrada y a la salida del receptor DS/BPSK en la figura 7.9. Con este fin recurrimos a la ecuación (7.13) para expresar la salida del detector coherente en la forma

$$\begin{aligned} v &= \sqrt{\frac{2}{T_b}}\int_0^{T_b} u(t)\,\cos(2\pi f_c t)\,dt \\ &= v_s + v_{cj} \end{aligned} \tag{7.23}$$

donde las componentes v_s y v_{cj} se deben a la señal PSK binaria desdispersada, $s(t)$, y la interferencia dispersa, $c(t)j(t)$, respectivamente. Estas dos componentes se definen del modo siguiente

$$v_s = \sqrt{\frac{2}{T_b}}\int_0^{T_b} s(t)\,\cos(2\pi f_c t)\,dt \tag{7.24}$$

y

$$v_{cj} = \sqrt{\frac{2}{T_b}}\int_0^{T_b} c(t)j(t)\,\cos(2\pi f_c t)\,dt \tag{7.25}$$

Considere primero el componente v_s debida a la señal. La señal PSK binaria desdispersada $s(t)$ es igual a

$$s(t) = \pm\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t), \quad 0 \le t \le T_b \tag{7.26}$$

donde el signo más corresponde al bit de información 1, y el signo menos al bit de información 0. De ese modo, suponiendo que la frecuencia de la portadora f_c es un múltiplo entero de $1/T_b$, tenemos

$$v_s = \pm\sqrt{E_b} \tag{7.27}$$

Considere a continuación la componente v_{cj} que resulta de la interferencia. Al expresar la señal de pseudo ruido $c(t)$ en la forma descrita de una secuencia $\{c_0, c_1, \ldots, c_{N-1}\}$, es posible reescribir la ecuación (7.25) en la forma correspondiente

$$v_{cj} = \sqrt{\frac{2}{T_b}} \sum_{k=0}^{N-1} c_k \int_{kT_c}^{(k+1)T_c} j(t) \cos(2\pi f_c t)\, dt \tag{7.28}$$

Al utilizar la ecuación (7.14) para $\phi_k(t)$, y luego la ecuación (7.18) para el coeficiente j_k, es posible redefinir v_{cj} como

$$\begin{aligned} v_{cj} &= \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} c_k \int_0^{T_b} j(t)\phi \cos(2\pi f_c t)\, dt \\ &= \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} c_k j_k \end{aligned} \tag{7.29}$$

A continuación aproximamos la secuencia de pseudo ruido como una *secuencia binaria e idénticamente distribuida* (*idd*). Subrayamos la implicación de esta aproximación reconstruyendo la ecuación (7.29) en la forma

$$V_{cj} \sqrt{\frac{T_c}{T_b}} \sum_{k=0}^{N-1} C_k j_k \tag{7.30}$$

donde V_{cj} y C_k son variables aleatorias con valores de muestreo v_{cj} y c_k, respectivamente. En la ecuación (7.30), el perturbador se supone fijo. Con las C_k tratadas como variables aleatorias idd, encontramos que la probabilidad del evento $C_k = \pm 1$ es

$$P(C_k = 1) = P(C_k = -1) = \tfrac{1}{2} \tag{7.31}$$

Por tanto, la media de variable aleatoria V_{cj} es cero, puesto que para k tenemos

$$\begin{aligned} E[C_k j_k | j_k] &= j_k P(C_k = 1) - j_k P(C_k = -1) \\ &= \tfrac{1}{2} j_k - \tfrac{1}{2} j_k \\ &= 0 \end{aligned} \tag{7.32}$$

Para un vector fijo $\mathbf{j}$, que representa el conjunto de coeficientes $j_0, j_1, \ldots, j_{N-1}$, la varianza de V_{cj} está dada por

$$\text{var}[V_{cj} | \mathbf{j}] = \frac{1}{N} \sum_{k=0}^{N-1} j_k^2 \tag{7.33}$$

Puesto que el *factor de dispersión* $N = T_b/T_c$, es viable utilizar la ecuación (7.22) para expresar esta varianza en términos de la potencia de interferencia promedio J como

$$\text{var}[V_{cj} | \mathbf{j}] = \frac{JT_c}{2} \tag{7.34}$$

De ese modo la variable aleatoria V_{cj} tiene media cero y varianza $JT_c/2$.

A partir de la ecuación (7.27) advertimos que la componente de la señal a la salida del detector coherente (durante cada intervalo de bit) es igual a $\pm\sqrt{E_b}$, donde E_b es la energía de la señal por bit. Por consiguiente, la potencia instantánea pico de la componente de la señal es E_b.

Por tanto, es posible definir la *relación señal de salida a ruido* como la potencia pico instantánea E_b dividida entre la varianza de la componente de ruido equivalente en la ecuación (7.34). Consecuentemente, escribimos

$$(\text{SNR})_O = \frac{2E_b}{JT_c} \tag{7.35}$$

La potencia de la señal promedio a la entrada del receptor es igual a E_b/T_b. De este modo definimos una relación *señal a ruido de entrada* como

$$(\text{SNR})_I = \frac{E_b/T_b}{J} \tag{7.36}$$

Por tanto, eliminando E_b/J entre las ecuaciones (7.35) y (7.36), podemos expresar la relación señal a ruido de salida en términos de la relación señal a ruido de entrada como

$$(\text{SNR})_O = \frac{2T_b}{T_c}(\text{SNR})_I \tag{7.37}$$

Es una práctica común expresar la relación señal a ruido en decibeles. Con ese propósito, introdujimos un término llamado la *ganancia de procesamiento* (PG), la cual se define como *la ganancia en la SNR obtenida mediante el uso del espectro disperso*. Específicamente, escribimos

$$\text{PG} = \frac{T_b}{T_c} \tag{7.38}$$

que representa la ganancia que se consigue al procesar una señal de espectro disperso sobre una señal no dispersa. Es posible escribir entonces la ecuación (7.37) en la forma equivalente:

$$10\log_{10}(\text{SNR})_O = 10\ \log_{10}(\text{SNR})_I + 3 + 10\ \log_{10}(\text{PG})\ \text{dB} \tag{7.39}$$

El término de 3 dB en el segundo miembro de la ecuación (7.39) corresponde a la ganancia en la SNR que se obtiene mediante el uso de detección coherente (lo cual presupone un conocimiento exacto de la fase de la señal por parte del receptor). Esta ganancia en la SNR no tiene nada que ver con el uso del espectro disperso. Más bien, es el último término, $10\ \log_{10}(\text{PG})$, el que toma en cuenta la ganancia de procesamiento. Note que tanto la ganancia de procesamiento PG como el factor de dispersión N (es decir, la longitud de la secuencia de pseudo ruido) igualan al cociente T_b/T_c. Por tanto, cuanto más grande hagamos la secuencia de pseudo ruido (o en forma correspondiente, sea más pequeño el tiempo de división T_c), tanto más grande será la ganancia de procesamiento.

7.6 *Probabilidad de error*

Considere que la salida v del detector coherente en el sistema BPSK de la figura 7.9 denota el valor de muestreo de una variable aleatoria V y que la componente de ruido equivalente v_{cj} producida por interferencia externa denota el valor de muestreo de una variable aleatoria V_{cj}. Entonces, a partir de las ecuaciones (7.23) y (7.27) deducimos que

$$V = \pm\sqrt{E_b} + V_{cj} \tag{7.40}$$

donde E_b es la energía de la señal transmitida por bit. El signo más se refiere a enviar el símbolo (bit de información) 1, y el signo menos a la transmisión del símbolo 0. La regla de decisión usada por el detector coherente de la figura 7.9 es declarar que el bit recibido en un intervalo $(0, T_b)$ es 1 si la salida del detector supera un umbral de 0, y que es 0 si la salida del detector es menor que el umbral. Si la salida del detector corresponde exactamente a 0, el receptor realiza

una suposición a favor de 1 o 0. Suponiendo igualmente probables ambos bits de información, encontramos que (debido a la naturaleza simétrica del problema) la probabilidad promedio de error P_e es la misma que la probabilidad condicional de que (digamos) el receptor tome una decisión a favor del símbolo 1, dado que se envió el símbolo 0. Esto es,

$$\begin{aligned} P_e &= P(V > 0 | \text{se envió el símbolo 0}) \\ &= P(V_{cj} > \sqrt{E_b}) \end{aligned} \tag{7.41}$$

En forma natural, la probabilidad de error P_e depende de la variable aleatoria V_{cj} definida por la ecuación (7.30). De acuerdo con esta definición, V_{cj} es la suma de N variables aleatorias distribuidas idénticamente. Por tanto, con base en el *teorema del límite central*, deducimos que para N grande la variable aleatoria V_{cj} asume una distribución gaussiana. En realidad, el factor de dispersión o la longitud N de la secuencia de pseudo ruido es por lo común grande en los sistemas de espectro disperso de secuencia directa que se encuentran en la práctica, bajo lo cual se justifica la condición de la aplicación del teorema de límite central.

Antes evaluamos la media y la varianza de V_{cj}; vea las ecuaciones (7.32) y (7.34). Por tanto, podemos establecer que la componente de ruido V_{cj} contenida en la salida del detector coherente puede aproximarse como una variable aleatoria gaussiana con media 0 y varianza $JT_c/2$, donde J es la potencia de interferencia promedio y T_c es la duración de la división. Con esta aproximación a la mano, podemos proceder entonces para calcular la probabilidad del evento $V_{cj} > \sqrt{E_b}$, y de este modo expresar la probabilidad promedio de error de acuerdo con la ecuación (7.41) como

$$P_e \simeq \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{JT_c}}\right) \tag{7.42}$$

Esta fórmula simple, la cual apela a la suposición gaussiana, es apropiada para sistemas binarios DS/BPSK con un factor de dispersión N.

■ CARACTERÍSTICAS DE ANTIPERTURBACIÓN

Resulta informativo comparar la ecuación (7.42) con la fórmula para la probabilidad promedio de error de un sistema PSK binario coherente que se reproduce aquí por conveniencia en la presentación [vea la ecuación (6.20)]

$$P_e = \frac{1}{2}\text{ferc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{7.43}$$

Con base en esta comparación, vemos que en cuanto a lo que tiene que ver con el cálculo de la tasa de error de bits en un sistema PSK binario disperso de secuencia directa, es posible considerar la interferencia como ruido de banda ancha de densidad espectral de potencia $N_0/2$, definida por

$$\frac{N_0}{2} = \frac{JT_c}{2} \tag{7.44}$$

Esta relación es simplemente un reenunciado del resultado anterior que se indicó en la ecuación (7.34).

Puesto que la energía de la señal por bit $E_b = PT_b$, donde P es la potencia promedio de la señal y T_b es la duración de bit, podemos expresar la relación de la energía de la señal por bit a la densidad espectral de ruido como

$$\frac{E_b}{N_0} = \left(\frac{T_b}{T_c}\right)\left(\frac{P}{J}\right) \tag{7.45}$$

Utilizando la definición de la ecuación (7.38) para la ganancia de procesamiento PG es posible reformular este resultado como

$$\frac{J}{P} = \frac{\text{PG}}{E_b / N_0} \tag{7.46}$$

El cociente J/P se denomina el *margen de perturbación*. Por tanto, este último y la ganancia de procesamiento, ambos expresados en decibeles, se relacionan por medio de

$$(\text{Margen de perturbación})_{\text{dB}} = (\text{Ganancia de procesamiento})_{\text{dB}} - 10 \log_{10}\left(\frac{E_b}{N_0}\right)_{\text{mín}} \tag{7.47}$$

donde $(E_b/N_o)_{\text{mín}}$ es el valor mínimo necesario para soportar una probabilidad promedio preestablecida de error.

▶ **Ejemplo 7.3**

Un sistema de comunicación de espectro disperso tiene los siguientes parámetros:

Duración del bit de información, $T_b = 4.095$ ms
Duración del recorte del pseudo ruido, $T_c = 1$ μs

Por tanto, utilizando la ecuación (7.38) encontramos que la ganancia de procesamiento es

$$\text{PG} = 4095$$

En forma correspondiente, el periodo requerido de la secuencia de pseudo ruido es $N = 4095$, y la longitud del registro de corrimiento es $m = 12$.

Para una recepción satisfactoria, suponemos que la probabilidad promedio de error no supera 10^{-5}. De la fórmula para un receptor PSK binario coherente, encontramos que $E_b/N_0 = 10$ produce una probabilidad promedio de error igual 0.387×10^{-5}. De ese modo, utilizando este valor de E_b/N_0, así como el valor calculado para la ganancia del procesamiento, encontramos con base en la ecuación (7.47) que el margen de perturbación es

$$\begin{aligned}(\text{Margen de perturbación})_{\text{dB}} &= 10 \log_{10} 4095 - 10 \log_{10}(10) \\ &= 36.1 - 10 \\ &= 26.1 \text{ dB}\end{aligned}$$

Esto es, los bits de información a la salida del receptor pueden detectarse en forma confiable incluso cuando el ruido o la interferencia a la entrada del receptor es hasta de 409.5 veces la potencia de la señal recibida. Claramente, ésta es una poderosa ventaja contra la interferencia (perturbación), la cual se pone en práctica a través del uso inteligente de la modulación por espectro disperso.

◀

7.7 *Espectro disperso de salto de frecuencia*

En el tipo de sistemas de espectro disperso que se presentaron en la sección 7.4, el uso de una secuencia de pseudo ruido para modular una señal de corrimiento de fase por llaveo logra una dispersión *instantánea* del ancho de banda de transmisión. La capacidad de un sistema de este tipo para combatir los efectos de los perturbadores se determina mediante la ganancia de procesamiento del sistema, que es una función del periodo de la secuencia de pseudo ruido. La ganancia de procesamiento puede crecerse empleando una secuencia de pseudo ruido con una estrecha duración de división, la cual, a su vez, permite un mayor ancho de banda de transmisión y más divisiones por bit. Sin embargo, las capacidades de los dispositivos físicos que se utilizan para generar la señal de espectro disperso de pseudo ruido imponen un límite práctico sobre la ganancia de procesamiento que puede obtenerse. En realidad, puede ocurrir que la ganancia de procesamiento que

se consigue de esa manera aún no sea lo suficientemente grande para superar los efectos de algunos perturbadores de interés, en cuyo caso tenemos que recurrir a otros métodos. Uno de tales métodos alternativos es obligar al perturbador a cubrir un espectro más ancho al *saltar al azar* la portadora modulada por los datos de una frecuencia a la siguiente. En efecto, el espectro de la señal transmitida se dispersa *en forma secuencial* en vez de instantáneamente; el término "secuencial" se refiere a la secuencia ordenada pseudoaleatoria de los saltos de frecuencia.

El tipo de espectro disperso en el cual la portadora salta aleatoriamente de una frecuencia a otra recibe el nombre de *espectro disperso de salto de frecuencia (FH)*. Un formato de modulación común para los sistemas FH es el de corrimiento de frecuencia por llaveo M-aria (MFSK). La combinación de estas dos técnicas se conoce simplemente como FH/MFSK. (Una descripción de la FSK M-aria se presenta en el capítulo 6.)

Puesto que el salto de frecuencia no cubre instantáneamente el espectro de dispersión completo, se nos obliga a considerar la tasa a la cual ocurren los saltos. En este contexto, es posible identificar dos caracterizaciones básicas (independientes de la tecnología) del salto de frecuencia:

1. *Salto de frecuencia lento*, en el cual la velocidad del símbolo R_s de la señal MFSK es un múltiplo entero de la *velocidad de salto* (R_h). Esto es, se transmiten varios símbolos en cada salto de frecuencia.
2. *Salto de frecuencia rápido*, en el cual la velocidad de salto R_h es un múltiplo entero de la velocidad del símbolo R_s de la MFSK. Esto es, la frecuencia de la portadora cambiará o saltará mediante la transmisión de un símbolo.

Evidentemente, el salto de frecuencia lento y el salto de frecuencia rápido son inversos uno del otro. En lo que sigue, las dos caracterizaciones de salto de frecuencia se consideran por separado.

■ SALTO DE FRECUENCIA LENTO

La figura 7.10*a* presenta un diagrama de bloques de un transmisor FH/MFSK, el cual incluye *modulación de frecuencia* seguida de *mezclado*. En primer lugar, los datos binarios entrantes se aplican a un modulador FSK M-ario. La onda modulada que se produce y la salida de un *sintetizador de frecuencia* digital se aplican entonces a un mezclador que consiste en un multiplicador seguido por un filtro pasobanda. Este último se diseña para elegir como señal transmitida la componente de frecuencia suma que resulta del proceso de multiplicación. En particular, segmentos sucesivos de k bits de una secuencia de pseudo ruido accionan al sintetizador de frecuencia, lo cual permite a la frecuencia de la portadora saltar a lo largo de 2^k valores distintos. En un solo salto, el ancho de banda de la señal transmitida es igual al que resulta del uso de una MFSK convencional con un alfabeto de $M = 2^K$ señales ortogonales. Sin embargo, para un intervalo completo de 2^k saltos de frecuencia, la señal FH/MFSK transmitida ocupa un ancho de banda mucho mayor. En realidad, son posibles los anchos de banda FH por encima de varios GHz, que es un orden de magnitud superior al que puede conseguirse con los espectros dispersos de secuencia directa. Una implicación de estos anchos de banda FH mayores es que la detección coherente es factible sólo dentro de cada salto, ya que los sintetizadores de frecuencia son incapaces de mantener la coherencia de fase a lo largo de saltos sucesivos. Por tanto, la mayor parte de los sistemas de espectro disperso y salto de frecuencia utilizan esquemas de modulación M-arios no coherentes.

En el receptor que se describe en la figura 7.10*b*, el salto de frecuencia se elimina primero mezclando (conversión descendente) la señal recibida con la salida del sintetizador de frecuencia local que se controla en forma síncrona, del mismo modo que en el transmisor. La salida resultante se aplica después a un filtro pasobanda y se procesa subsecuentemente mediante un detector FSK M-ario *no coherente*. Para realizar este detector M-ario, podemos utilizar un banco de M filtros de acoplamiento no coherentes, cada uno de los cuales se acopla a uno de los tonos MFSK. (Los filtros de acoplamiento no coherentes se describen en el capítulo 6.) Una estimación del símbolo original transmitido se obtiene al elegir la salida de filtro más grande.

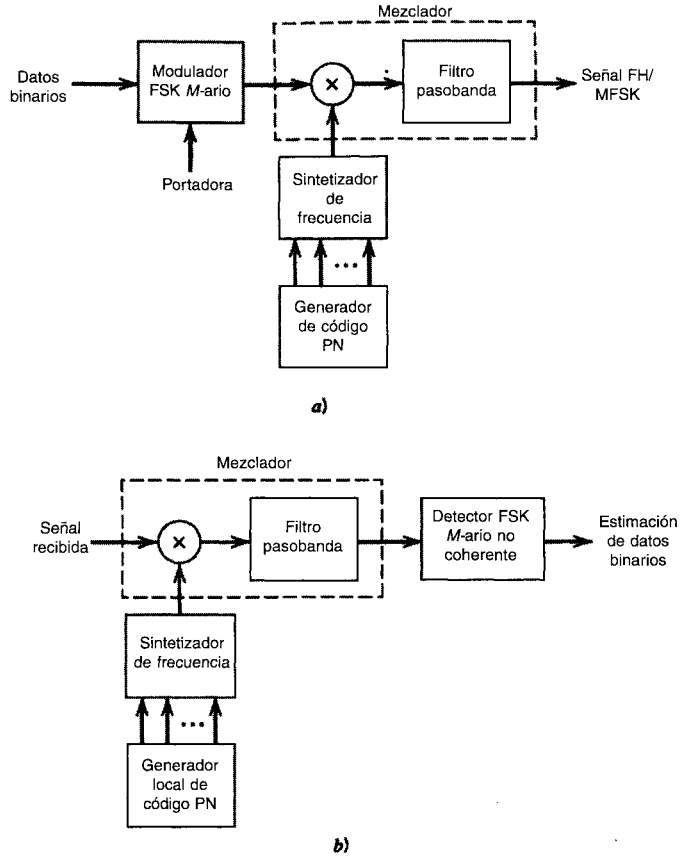


FIGURA 7.10 Corrimiento de frecuencia por llaveo M-ario disperso con salto de este mismo. *a*) Transmisor. *b*) Receptor.

Un tono FH/MFSK individual de duración más corta se conoce como una *división*; esta terminología no debe confundirse con la que se usa en la sección 7.4 al describir la DS/BPSK. La *tasa de división*, R_c, para un sistema FH/MFSK está definida por

$$R_c = \text{máx}(R_h, R_s) \tag{7.48}$$

donde R_h es la *tasa de salto* y R_s es la *tasa de símbolos*.

Una señal FH/MFSK lenta se caracteriza por tener símbolos múltiples transmitidos por salto. Por tanto, cada símbolo de una señal FH/MFSK es una división. De modo correspondiente, en un sistema FH/MFSK lento, la tasa de bits R_b de los datos binarios entrantes, la tasa de símbolo R_s de la señal MFSK, la tasa de división R_c y la tasa de salto R_h se relacionan por medio de

$$R_c = R_s = \frac{R_b}{K} \geq R_h \tag{7.49}$$

donde $K = \log_2 M$.

En cada salto, los tonos MFSK están separados en frecuencia por un múltiplo de la tasa de división $R_c = R_s$, lo que asegura su ortogonalidad. La implicación de esta condición es que todo símbolo transmitido no producirá ninguna diafonía en ninguno de los otros $M - 1$ filtros de acoplamiento no coherentes que constituyen el detector MFSK del receptor en la figura 7.10*b*. Por "diafonía" entendemos el derramamiento de la salida de un filtro en uno adyacente. El desempeño resultante del sistema FH/MFSK lento es igual al de la detección no coherente de

señales MFSK convencionales (sin salto) en ruido blanco gaussiano aditivo. De ese modo, la señal de interferencia (perturbadora) tiene un efecto sobre el receptor FH/MFSK, en términos de la probabilidad promedio de error del símbolo equivalente a la del ruido gaussiano blanco aditivo sobre un receptor FSK M-ario no coherente convencional que no experimenta interferencia. Con base en esta equivalencia, es posible utilizar la ecuación (6.140) para aproximar la evaluación de la probabilidad del error de símbolo en el sistema FH/MFSK.

Suponiendo que el perturbador decide distribuir su potencia promedio J por el espectro completo de frecuencias saltadas, el efecto perturbador es equivalente a un AWGN con densidad espectral de potencia $N_0/2$, donde $N_0 = J/W_c$ y W_c es el ancho de banda FH. El sistema de espectro disperso está caracterizado, de tal modo, por la relación de energía del símbolo a la densidad espectral del ruido:

$$\frac{E}{N_0} = \frac{P/J}{W_c/R_s} \tag{7.50}$$

donde la proporción P/J es el recíproco del margen de perturbación. El otro cociente en el denominador de la ecuación (7.50) es la ganancia de procesamiento del sistema FH/MFSK lento, la cual está definida por

$$\begin{aligned} \text{PG} &= \frac{W_c}{R_s} \\ &= 2^k \end{aligned} \tag{7.51}$$

Esto es, la ganancia de procesamiento (expresada en decibeles) es igual a $10 \log_{10} 2^k \simeq 3k$, donde k es la longitud del segmento de pseudo ruido empleado para elegir un salto de frecuencia.

Este resultado supone que el perturbador distribuye su potencia por el espectro FH completo. Sin embargo, si el perturbador decide concentrarse en unas cuantas frecuencias de salto, entonces la ganancia de procesamiento producida por el receptor sería menor que $3k$ decibles.

▶ EJEMPLO 7.4

La figura 7.11*a* ilustra la variación de la frecuencia de una señal FH/MFSK lenta con el tiempo para un periodo completo de la secuencia de pseudo ruido. El periodo de la frecuencia de pseudo ruido corresponde a $2^4 - 1 = 15$. La señal FH/MFSK tiene los siguientes parámetros:

Número de bits por símbolo MFSK	$K = 2$
Número de tonos MFSK	$M = 2^K = 4$
Longitud del segmento de pseudo ruido por salto	$k = 3$
Número total de saltos de frecuencia	$2^k = 8$

En este ejemplo, la portadora salta a una nueva frecuencia después de transmitir dos símbolos o, equivalentemente, cuatro bits de información. La figura 7.11*a* incluye los datos binarios de entrada, así como la secuencia de pseudo ruido que controla la selección de la frecuencia de la portadora FH. Vale la pena señalar que aunque hay ocho frecuencias disponibles para el salto, sólo tres de ellas las utiliza la secuencia de pseudo ruido.

La figura 7.11*b* presenta la variación con el tiempo de la frecuencia a la que se le ha suprimido el salto. Se reconoce que esta variación es la misma que la de la señal MFSK convencional producida por los datos de entrada dados. ◀

■ SALTO DE FRECUENCIA RÁPIDO

Un sistema FH/MFSK rápido difiere de un sistema FH/MFSK lento en que hay saltos múltiples por símbolo M-ario. Por consiguiente, en el sistema rápido señalado cada salto es una división.

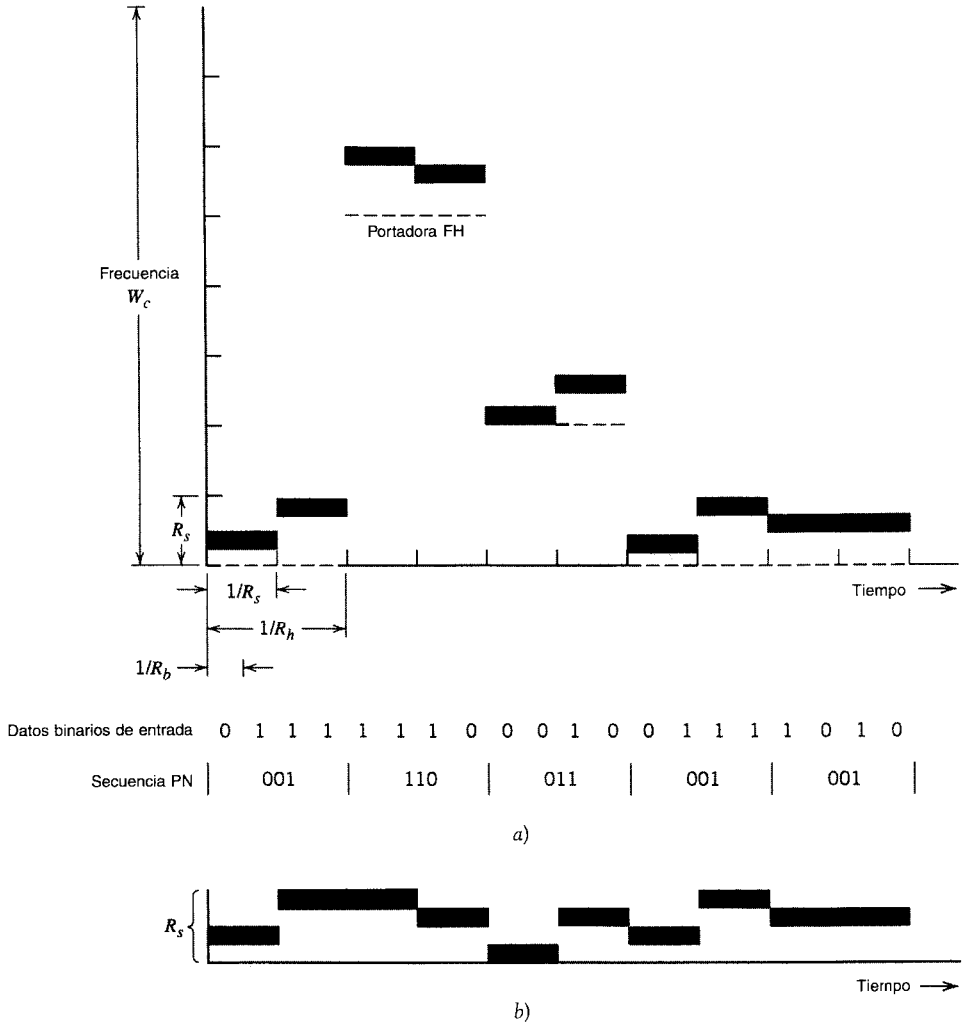


FIGURA 7.11 Ilustración del salto de frecuencia lento. *a)* Variación de frecuencia para un periodo completo de la secuencia PN. *b)* Variación de la frecuencia sin salto con el tiempo.

En general, el salto de frecuencia rápido se usa para vencer a una táctica de perturbación inteligente que implica dos funciones: mediciones de la señal transmitida y resintonización de la señal de interferencia con respecto a esa porción de la banda de frecuencia. Claramente, para vencer al perturbador, la señal transmitida debe saltar a una nueva frecuencia antes de que el perturbador sea capaz de completar el procesamiento de estas dos funciones.

Para la recuperación de datos en el receptor se usa la detección no coherente. Sin embargo, el procedimiento de detección es bastante diferente del que se usa en el receptor FH/MFSK lento. En particular es posible considerar dos procedimientos:

1. Para cada símbolo FH/MFSK se realizan decisiones independientes de las K divisiones de salto de frecuencia recibidas, y una regla simple basada en un *voto de mayoría* se usa para efectuar una estimación del símbolo MFSK con el salto eliminado.
2. Para cada símbolo FH/MFSK, las funciones de verosimilitud se calculan como funciones de la señal total recibidas sobre K divisiones, y se elige la más grande.

Un receptor basado en el segundo procedimiento es óptimo porque minimiza la probabilidad promedio del error de símbolo para una E_b/N_0 dada.

▶ EJEMPLO 7.5

La figura 7.12*a* ilustra la variación en el tiempo de la frecuencia transmitida de una señal FH/MFSK rápida. La señal tiene los siguientes parámetros:

Número de bits por símbolo MFSK	$K = 2$
Número de tonos MFSK	$M = 2^K = 4$
Longitud del segmento de pseudo ruido por salto	$k = 3$
Número total de saltos de frecuencia	$2^k = 8$

En este ejemplo cada símbolo MFSK tiene el mismo número de bits y divisiones; es decir, la tasa de división R_c es la misma que la tasa de bits R_b. Después de cada división, la frecuencia de portadora de la señal MFSK transmitida salta a un valor diferente, excepto en unas cuantas ocasiones cuando el segmento de k divisiones de una secuencia de pseudo ruido se repite a sí mismo.

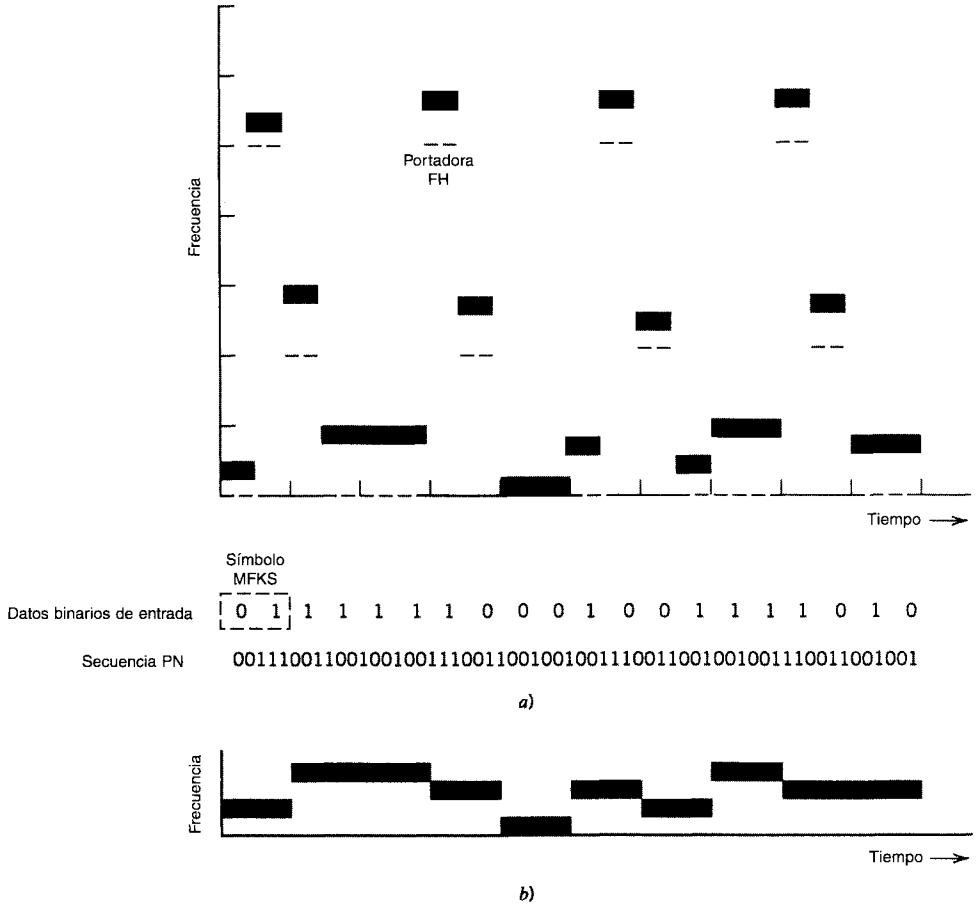


FIGURA 7.12 Ilustración de salto de frecuencia rápido. *a*) Variación de la frecuencia del transmisor con el tiempo. *b*) Variación de la frecuencia sin salto con el tiempo.

La figura 7.12*b* describe la variación en el tiempo de la frecuencia de una señal MFSK en la que se ha suprimido el salto, la cual es la misma que la del ejemplo 7.4. ◀

7.8 *Experimentos de computadora: longitud máxima y códigos de Gold*

El *multiplexado por división de código* (CDM) proporciona una alternativa a los métodos tradicionales de multiplexado por división de frecuencia (FDM) y multiplexado por división de tiempo (TDM). No requiere la asignación del ancho de banda del FDM (explicado en el capítulo 2) ni la sincronización de tiempo necesaria en el TDM (analizado en el capítulo 3). En vez de eso, a los usuarios de un canal común se les permite el acceso a éste a través de la asignación de un "código disperso" para cada usuario individual bajo el amparo de la modulación por espectro disperso. El propósito de este experimento de computadora es estudiar ciertas clases de códigos de dispersión para sistemas CDM que ofrecen un desempeño satisfactorio.

En un sistema CDM ideal, la relación cruzada entre cualesquiera dos usuarios del sistema es cero. Para que se produzca esta condición ideal, requerimos que la función de correlación cruzada entre los códigos de dispersión asignados a cualesquiera dos usuarios del sistema sea cero para todos los corrimientos cíclicos. Desafortunadamente, las secuencias de pseudo ruido ordinarias no satisfacen este requerimiento debido a sus propiedades de correlación de cruce relativamente pobres.

Como remedio para esta desventaja de las secuencias de pseudo ruido binarias, es factible utilizar una clase especial de secuencias de pseudo ruido conocidas como *secuencias (códigos) Gold*,[5] cuya generación se encuentra incorporada en el siguiente teorema:

> Sean $g_1(X)$ y $g_2(X)$ un par preferido de polinomios primitivos de grado n cuyos registros de corrimiento correspondientes generan secuencias de longitud máxima de periodo $2^n - 1$ y cuya función de correlación cruzada tiene una magnitud menor o igual que
>
> $$2^{(n+1)/2} + 1 \quad \text{para } n \text{ impar} \tag{7.52}$$
>
> o
>
> $$2^{(n+2)/2} + 1 \quad \text{para } n \text{ par y } n \neq 0 \bmod 4 \tag{7.53}$$
>
> En ese caso el registro de corrimiento correspondiente al polinomio del producto $g_1(X) \cdot g_2(X)$ generará $2^n + 1$ secuencias diferentes, donde cada secuencia tendrá un periodo de $2^n - 1$, y la correlación cruzada entre cualquier par de dichas frecuencias satisface la condición anterior.

De aquí en adelante este teorema se denominará *teorema de Gold*.

Para comprender el teorema de Gold necesitamos definir lo que entendemos por un polinomio primitivo. Considere un polinomio $g(X)$ definido sobre un *campo binario* (es decir, un conjunto finito de dos elementos, 0 y 1, los cuales están gobernados por las reglas de la aritmética binaria). Se dice que el polinomio $g(X)$ será un *polinomio irreductible* si no puede factorizarse utilizando cualesquiera polinomios del campo binario. Un polinomio irreductible $g(X)$ se dice que será un *polinomio primitivo* si el entero m más pequeño para el cual el polinomio $g(X)$ divide al factor $X^n + 1$ es $n = 2^m - 1$. Una explicación adicional de este tema se posterga para el capítulo 8; en particular vea el ejemplo 8.3.

Experimento 1. Propiedades de correlación de las secuencias de pseudo ruido

Considere un par de registros de corrimiento para generar dos secuencias de pseudo ruido del periodo $2^7 - 1 = 127$. Un registro de corrimiento realimentado tiene las derivaciones de realimentación [7, 1], y el otro, las derivaciones de realimentación [7, 6 , 5, 4]. Ambas secuencias tienen la misma función de autocorrelación que se muestra en la figura 7.13*a*, la cual surge de inmediato de la definición que se presentó en la ecuación (7.5).

Sin embargo, el cálculo de la función de correlación cruzada entre las secuencias de pseudo ruido es una proposición más difícil, particularmente para n grande. A fin de realizar este cálculo, recurrimos al uso de la simulación por computadora para el corrimiento cíclico variable τ dentro del intervalo $0 < \tau \leq 2^n - 1$. Los resultados de este cálculo se presentan en la figura 7.13*b*. Esta última confirma la pobre propiedad de correlación cruzada de las secuencias de pseudo ruido en comparación con su función de autocorrelación. La magnitud de correlación cruzada excede 40.

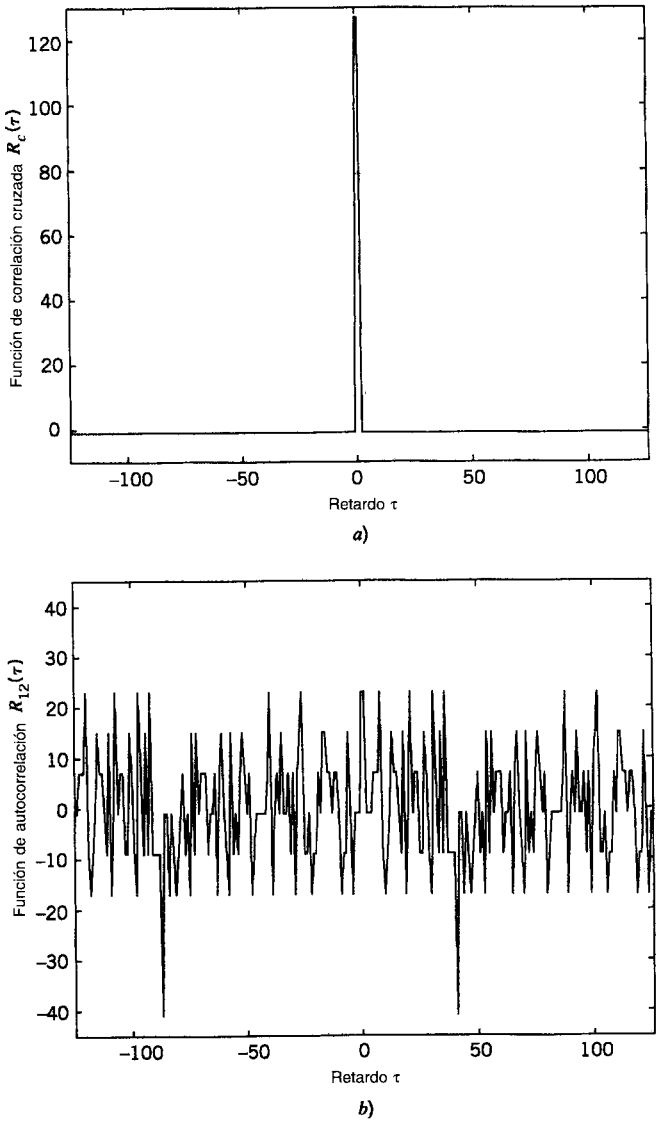


FIGURA 7.13 *a*) Función de autocorrelación $R_c(\tau)$, y *b*) función de correlación cruzada $R_{12}(\tau)$ de las dos secuencias PN [7, 1] y [7, 6, 5, 4].

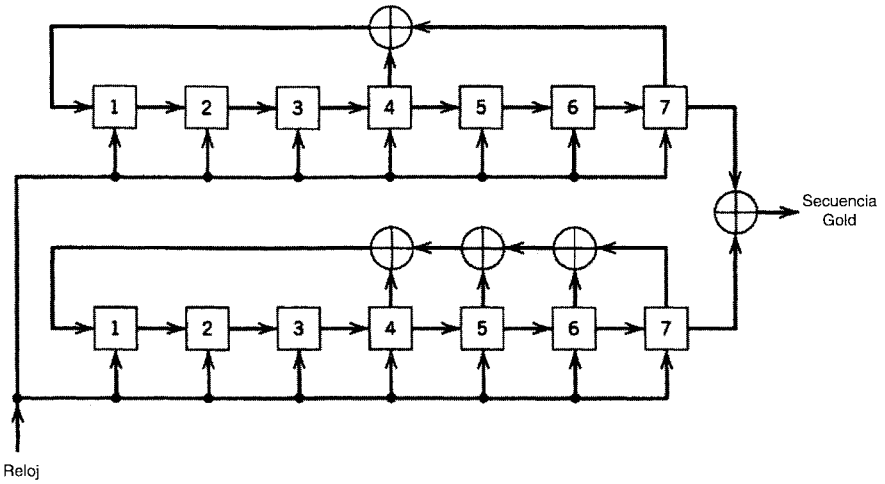


FIGURA 7.14 Generador para una secuencia Gold de periodo $2^7 - 1 = 127$.

Experimento 2. Propiedades de correlación de las secuencias Gold

Para nuestro siguiente experimento, consideramos secuencias Gold con periodo $2^7 - 1 = 127$. Para generar secuencias de este tipo correspondientes a $n = 7$, necesitamos un par preferencial de secuencias de pseudo ruido que satisfagan la ecuación (7.52) (n impar), según indica

$$2^{(n+1)/2} + 1 = 2^4 + 1 = 17$$

Este requerimiento se satisface mediante las secuencias de pseudo ruido con derivaciones de realimentación [7, 4] y [7, 6, 5, 4]. El generador de la secuencia Gold se presenta en la figura 7.14 e incluye la adición de módulo 2 de estas dos secuencias. De acuerdo con el teorema de Gold hay un total de

$$2^n + 1 = 2^7 + 1 = 129$$

secuencias que cumplen la ecuación (7.52). La correlación entre cualquier par de tales secuencias se muestra en la figura 7.15, la cual concuerda plenamente con el teorema de Gold. En particular, la magnitud de la correlación cruzada es menor o igual a 17.

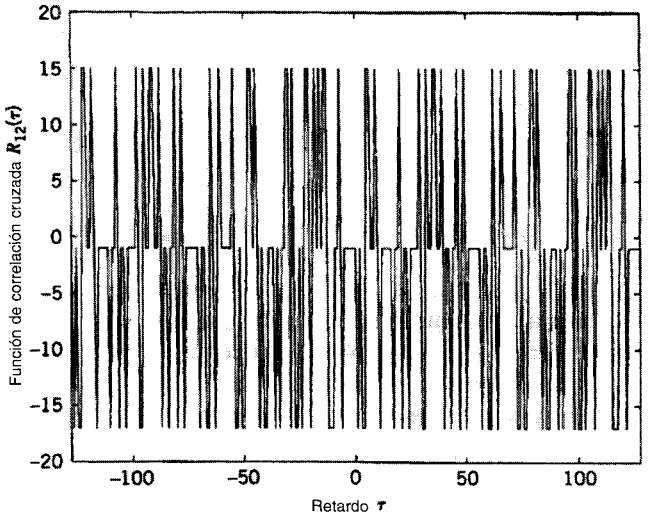


FIGURA 7.15 Función de correlación cruzada $R_{12}(\tau)$ de un par de secuencias de Gold basada en las dos secuencias PN [7, 4] y [7, 6, 5, 4].

7.9 Resumen y discusión

El *corrimiento de fase por llaveo M-aria de secuencia directa* (DS/MPSK) y el *corrimiento de frecuencia por llaveo M-aria de salto de frecuencia* (FH/MFSK) representan dos categorías principales de las comunicaciones de espectro disperso. Ambas se basan en el uso de la secuencia de pseudo ruido, la cual se aplica directamente en dos categorías.

En un sistema DS/MPSK, la secuencia de pseudo ruido hace que la señal transmitida asuma una apariencia similar al ruido al dispersar su espectro simultáneamente sobre un intervalo amplio de frecuencias. En el corrimiento de fase por llaveo es posible utilizar la PSK binaria (es decir, $M = 2$) con una portadora simple. De modo alternativo, podemos utilizar la QPSK (es decir, $M = 4$), en cuyo caso los datos se transmiten utilizando un par de portadoras en cuadratura de fase. (Tanto la PSK como la QPSK se explican en la sección 6.3.) La motivación usual para utilizar la QPSK es proporcionar una eficiencia del ancho de banda mejorada. En un sistema de espectro disperso, la eficiencia del ancho de banda no suele ser de interés fundamental. En vez de eso, el uso de la QPSK está motivado por el hecho de que ésta es menos sensible a algunos tipos de interferencia (perturbación).

En un sistema FH/MFSK, la secuencia de pseudo ruido realiza el salto de la portadora sobre varias frecuencias de una manera pseudoaleatoria, con el resultado de que el espectro de la señal transmitida se dispersa de modo secuencial.

Naturalmente, las técnicas de dispersión de espectro de secuencia directa y de salto de frecuencia pueden emplearse en un sistema simple. El sistema que resulta se conoce como *sistema de espectro disperso DS/FH híbrido*. La razón por la que se busca un método híbrido radica en que las ventajas de ambas técnicas de dispersión de espectro de secuencia directa y de salto de frecuencia se ponen en práctica en el mismo sistema.

Una exposición de las comunicaciones de espectro disperso sería incompleta sin alguna referencia a las formas de onda del perturbador. Los perturbadores que se encuentran en la práctica incluyen los siguientes tipos:

1. *El perturbador de ruido de barrera*, el cual consiste en ruido blanco gaussiano limitado en banda de alta potencia promedio. Éste es un perturbador de fuerza bruta que no explota ningún conocimiento del sistema de comunicación antiperturbación, con excepción de su ancho de banda disperso.
2. *El perturbador de ruido de banda parcial*, que consiste en ruido cuya potencia total se distribuye uniformemente sobre alguna banda de frecuencia que es un subconjunto del ancho de banda de dispersión total. Debido al ancho de banda más pequeño, este perturbador es más fácil de generar que el de ruido de barrera.
3. *El perturbador de ruido pulsado*, que implica la transmisión de ruido de banda ancha de potencia

$$J_{\text{pico}} = \frac{J}{p}$$

durante una fracción p del tiempo, y nada para la fracción de tiempo restante $1 - p$. La potencia promedio de ruido es igual a J.
4. *El perturbador de un solo tono*, que consiste en una onda senoidal cuya frecuencia se encuentra dentro del ancho de banda disperso; como tal, es la más fácil de generar de todas las señales perturbadoras.
5. *El perturbador multitonos*, que es el equivalente de tonos de un perturbador de ruido de banda parcial.

Aparte de estas cinco, muchos otros tipos de forma de onda de perturbación ocurren en la práctica. En cualquier caso, no hay una forma de onda perturbadora sencilla que sea la peor

para todos los sistemas de espectro disperso, y no hay un sistema de espectro disperso sencillo que sea el mejor contra todas las formas de onda de perturbación posibles.

NOTAS Y BIBLIOGRAFÍA

1. La definición de la modulación por espectro disperso presentada en la introducción se adaptó de Pickholtz, Schilling y Milstein [342]. Este artículo presenta un instructivo repaso de las teorías de las comunicaciones por espectro disperso.

   En lo referente a artículos introductorios sobre el tema, vea Viterbi [390] y Cook y Marsh [222]. En lo que se refiere a libros, vea Dixon [36], Holmes [69], Ziemer y Peterson ([177], pp. 327-649), Cooper y McGillem (pp. 269-411), y Simon, Omura, Scholtz y Levitt ([139], volúmenes I, II y III). La obra de tres volúmenes de Simon y otros es el tratamiento más amplio de las comunicaciones por espectro disperso disponible en la literatura pública. El desarrollo de las comunicaciones por espectro disperso se remonta a la mitad de la década de 1950. Con respecto a una explicación histórica de estas técnicas, vea Sholz [366]. Este último artículo describe los orígenes de las comunicaciones por espectro disperso desde la década de 1920. Gran parte del material histórico presentado en este artículo se reproduce en el capítulo 2, volumen 1, del libro de Simon y otros.

   El libro editado por Tantaratana y Ahmed [153] incluye artículos introductorios y avanzados sobre las aplicaciones inalámbricas de la modulación por espectro disperso. Los artículos se agrupan en las siguientes categorías: tecnología de espectro disperso, sistemas móviles celulares, comunicaciones de satélite, redes de área local inalámbricas y sistema de posición geográfica (PGS).

2. Para mayores detalles acerca de las secuencias de longitud máxima, consulte Golomb ([54], pp. 1-32), Simon, Omura, Scholtz y Levitt ([139], pp. 283-295), y Peterson y Weldon [118]. La última referencia incluye una lista amplia de polinomios para generar secuencias de longitud máxima; vea también Dixon [36]. Con respecto a un artículo instructivo acerca de las secuencias de pseudo ruido, vea Sarwate y Pursley [363].

3. La tabla 7.1 se extrajo del libro de Dixon ([36], pp. 81-83), donde las conexiones de realimentación de las secuencias de longitud máxima se tabulan para la longitud m del registro de corrimiento extendiéndose hasta 89.

4. En cuanto a una explicación detallada del problema de sincronización en las comunicaciones por espectro disperso, vea Ziemer y Peterson ([177], capítulos 9 y 10) y Simon y otros ([139], volumen III).

5. Los artículos originales acerca de las secuencias de Gold se encuentran en Gold [265, 266]. Una explicación amplia de las secuencias de Gold se presenta en Holmes [69].

PROBLEMAS

Secuencias de pseudo ruido

7.1 Una secuencia de pseudo ruido se genera utilizando un registro de corrimiento realimentado de longitud $m = 4$. La tasa de división es de 10^7 divisiones por segundo. Encuentre los siguientes parámetros:

a) Longitud de la secuencia de pseudo ruido.

b) Duración de la división de la secuencia de pseudo ruido.

c) Periodo de la secuencia de pseudo ruido.

7.2 La figura P7.2 muestra un registro de corrimiento de cuatro etapas. El estado inicial del registro es 1000. Determine la secuencia de salida del registro de corrimiento.

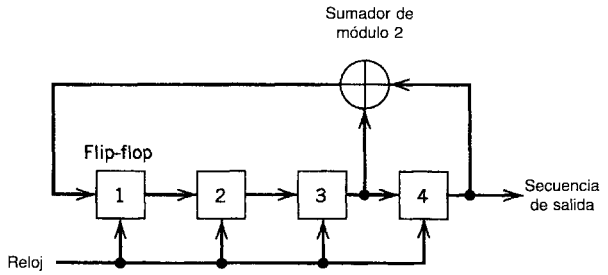


FIGURA P7.2

7.3 Para el registro de corrimiento realimentado dado en el problema 7.2, demuestre la propiedad de balance y la propiedad de ejecución de una secuencia de pseudo ruido. Además, calcule y grafique la función de autocorrelación de la secuencia de pseudo ruido producida por este registro de corrimiento.

7.4 En referencia a la tabla 7.1, elabore códigos de longitud máxima para las tres configuraciones de realimentación [6, 1], [6, 5, 2, 1] y [6, 5, 3, 2], cuyo periodo es $N = 63$.

7.5 La figura P7.5 muestra la versión multiderivación modular del registro de corrimiento realimentado lineal que se muestra en la figura 7.4*b*. Demuestre que la secuencia de pseudo ruido generada por este esquema es exactamente la misma que la descrita en la tabla 7.2*b*.

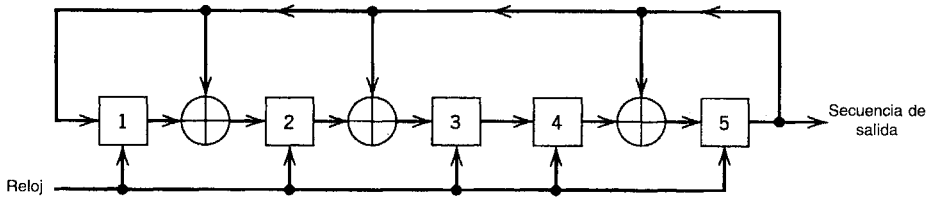


FIGURA P7.5

Sistema de secuencia directa/corrimiento de fase por llaveo

7.6 Demuestre que la tabla de verdad que se presenta en la tabla 7.3 puede construirse combinando los dos pasos siguientes:

a) Se aplica la suma módulo 2 a la señal de mensaje $b(t)$ y la señal de pseudo ruido $c(t)$.

b) Los símbolos 0 y 1 a la salida del sumador módulo 2 se representan mediante corrimientos de 0 y 180 grados, respectivamente.

7.7 Un perturbador de un solo tono

$$j(t) = \sqrt{2J}\cos(2\pi f_c t + \theta)$$

se aplica a un sistema DS/BPSK. La señal N-dimensional transmitida $x(t)$ se describe mediante la ecuación (7.16). Determine las $2N$ coordenadas del perturbador $j(t)$.

7.8 La ganancia de procesamiento de un sistema de espectro disperso puede expresarse como el cociente entre el ancho de banda disperso de la señal transmitida y el ancho de banda desdisperso de la señal recibida. Justifique este enunciado para el sistema DS/BPSK.

7.9 Un sistema de corrimiento de fase por llaveo binaria dispersa de secuencia directa utiliza un registro de corrimiento realimentado de longitud 19 para la generación de la secuencia de pseudo ruido. Calcule la ganancia de procesamiento del sistema.

7.10 En un sistema DS/BPSK, el registro de corrimiento realimentado que se utiliza para generar la secuencia de pseudo ruido tiene longitud $m = 19$. Se requiere que el sistema tenga una probabilidad promedio del error de símbolo debida a las señales de interferencia generadas externamente que no sea mayor que 10^{-5}. Calcule los siguientes parámetros del sistema en decibeles:

a) Ganancia del procesamiento.

b) Margen de antiperturbación.

7.11 En la sección 7.5 presentamos un análisis acerca de la dimensionalidad del espacio de señales y de la ganancia de procesamiento de un sistema de espectro disperso de secuencia directa, utilizando corrimiento de fase por llaveo binaria. Extienda el análisis que se presentó ahí al caso de un sistema de ese tipo que utiliza corrimiento de cuadrifase por llaveo.

Espectro disperso de salto de frecuencia

7.12 Un sistema FH/MFSK lento tiene los siguientes parámetros:

Número de bits por símbolo MFSK = 4

Número de símbolos MFSK por salto = 5

Calcule la ganancia de procesamiento del sistema.

7.13 Un sistema FH/MFSK rápido tiene los siguientes parámetros:

Número de bits por símbolo MFSK = 4

Número de saltos por símbolo MFSK = 4

Calcule la ganancia de procesamiento del sistema.

Experimentos de computadora

7.14 Considere dos secuencias de pseudo ruido $N = 63$. Una secuencia tiene las derivaciones de realimentación [6, 1] y la otra las derivaciones de realimentación [6, 5, 2, 1], las cuales se eligen de acuerdo con la tabla 7.1.

a) Calcule la función de autocorrelación de estas dos secuencias, y su función de correlación cruzada.

b) Compare la función de correlación cruzada calculada en la parte *a* con la función de correlación cruzada entre la secuencia [6, 5, 2, 1] y su imagen de espejo [6, 5, 4, 1].

Comente sus resultados.

7.15 **a)** Calcule la función de correlación cruzada de una secuencia de pseudo ruido con derivaciones de realimentación [5, 2] y su secuencia imagen definida por las derivaciones de realimentación [5, 3].

b) Repita el cálculo de la secuencia de pseudo ruido con derivaciones realimentadas [5, 2] y la secuencia de pseudo ruido con derivaciones de realimentación [5, 4, 2, 1].

c) Repita el cálculo de la secuencia de pseudo ruido con derivaciones de realimentación [5, 4, 3, 2] y la secuencia de pseudo ruido con derivaciones de realimentación [5, 4, 2, 1].

Las derivaciones de realimentación [5, 2], [5, 4, 3, 2] y [5, 4, 2, 1] son derivaciones posibles para una secuencia de longitud máxima de periodo 31, de acuerdo con la tabla 7.1.

CAPÍTULO 8

COMUNICACIONES DE RADIO MULTIUSUARIO

Como su nombre implica, las comunicaciones multiusuario se refieren al uso simultáneo de un canal de comunicaciones por parte de varios usuarios. En este capítulo presentamos los sistemas de comunicación multiusuario, que se fundamentan en la propagación del radio para enlazar los receptores con los transmisores.

En particular, nos centramos en los siguientes temas:

- *Técnicas de acceso múltiple, las cuales resultan básicas para los sistemas de comunicación multiusuario.*
- *Comunicaciones por satélite, que ofrecen cobertura mundial.*
- *Análisis del enlace de radio, subrayando los papeles de las antenas de transmisión y recepción y de la propagación por el espacio libre.*
- *Comunicaciones inalámbricas, destacando la movilidad y el fenómeno de multitrayectoria.*
- *La codificación del habla para las comunicaciones inalámbricas.*
- *Antenas adaptables para las comunicaciones inalámbricas.*

8.1 *Introducción*

Gran parte del material acerca de la teoría de comunicaciones que se ha presentado en los capítulos anteriores se ha basado en la idealización particular del canal de comunicación, a saber, un *modelo de canal limitado en ancho de banda y corrompido por ruido blanco gaussiano aditivo* (AWGN). La *teoría clásica de la comunicación* desarrollada de ese modo es matemáticamente elegante, y ofrece una sólida introducción al campo siempre en crecimiento de los sistemas de comunicaciones. Un ejemplo de un canal físico que es bien representado por un modelo de este tipo es el canal de comunicaciones por satélite. Por tanto, resulta conveniente que el primer tipo de comunicaciones multiusuario que se presenta en este capítulo corresponda a las *comunicaciones por satélite*.

Un sistema de comunicación por satélite en órbita geoestacionaria se basa en la propagación por radio de la línea de visión para la operación de su enlace ascendente desde una terminal terrestre hasta el transpondedor y el enlace descendente desde este último hasta otra terminal terrestre. De ese modo, la explicación de las comunicaciones por satélite conduce naturalmente al análisis de la propagación de las ondas de radio en el espacio libre, enlazando una antena receptora con una antena transmisora.

El uso de las comunicaciones por satélite ofrece *cobertura mundial*. El otro sistema de comunicación multiusuario que se estudia en este capítulo, es decir, las *comunicaciones inalámbricas*, ofrece la *movilidad* que, junto con las redes telefónicas existentes y los sistemas de comunicación por satélite, permiten que una unidad móvil se comunique con cualquiera y en cualquier parte en el mundo. Otro rasgo característico de los sistemas de comunicación

inalámbricos es que no *implican trabas* (es decir, se permite la total libertad de localización), y por ello el interés en su uso en redes de área local (es decir, redes de datos confinados a instalaciones hasta de unos cuantos kilómetros de tamaño), debido a las ventajas importantes sobre el cableado convencional: eliminación de alambrado y realambrado, flexibilidad en la creación de nuevos servicios de comunicación y movilidad de los usuarios.

El canal de propagación por ondas de radio que caracteriza a las comunicaciones inalámbricas se aparta del modelo de canal idealizado a AWGN debido a la presencia de *multitrayectorias*, que constituye una forma no gaussiana del fenómeno dependiente de la señal que surge por las reflexiones de la señal transmitida a partir de objetos fijos y en movimiento. La presencia de la multitrayectoria da lugar a dificultades prácticas en el uso del canal de propagación de radio y complica su análisis matemático. En términos simples, la multitrayectoria es un fenómeno físico intrínseco a la operación de las formas en interiores y en exteriores de las comunicaciones inalámbricas.

Sin embargo, antes de proceder a explicar aspectos específicos de las comunicaciones por satélite e inalámbricas es apropiado iniciar la exposición describiendo técnicas de acceso múltiple, las cuales permiten acceso simultáneo (o casi) a un canal común a diferentes usuarios.

8.2 *Técnicas de acceso múltiple*

El *acceso múltiple* es una técnica por medio de la cual muchos suscriptores o estaciones locales pueden compartir el uso de un canal de comunicación al mismo tiempo o casi de esa forma, a pesar de que sus transmisiones individuales puedan originarse desde ubicaciones bastante diferentes. Establecido de otra manera, una técnica de acceso múltiple permite que un gran número de usuarios que buscan comunicarse entre sí compartan los recursos de comunicación del canal.

Existen diferencias sutiles entre el acceso múltiple y el multiplexado que deben señalarse:

- El acceso múltiple se refiere a que usuarios en localidades altamente dispersas compartan en forma remota un canal de comunicación, como un satélite o un canal de radio. Por otra parte, el multiplexado se refiere a que usuarios confinados a un sitio local compartan un canal, como el telefónico.
- En un sistema multiplexado, los requerimientos del usuario son por lo general fijos. En contraste, en un sistema de acceso múltiple dichos requerimientos pueden cambiar en forma dinámica con el tiempo, en cuyo caso son necesarias provisiones para la asignación dinámica del canal.

Por razones evidentes es deseable que en un sistema de acceso múltiple se logre compartir los recursos del canal sin provocar interferencia seria entre los usuarios del sistema. En este contexto, es posible identificar cuatro tipos básicos de acceso múltiple:

1. *Acceso múltiple por división de frecuencia (FDMA).*
En esta técnica se asignan sub-bandas disjuntas de frecuencias a diferentes usuarios en un esquema de tiempo continuo. Para reducir la interferencia entre usuarios asignados a bandas de canal adyacente, se utilizan *bandas de seguridad*, para actuar como zonas de búfer, del modo que se ilustra en la figura 8.1*a*. Estas bandas de seguridad resultan necesarias debido a la imposibilidad de conseguir filtrado ideal para separar a los diferentes usuarios.

2. *Acceso múltiple por división del tiempo (TDMA).*
En esta segunda técnica, a cada usuario se le asigna la ocupación espectral completa del canal, pero sólo por una corta duración llamada *franja de tiempo*. Como se muestra en la figura 8.1*b*, las zonas de búfer en la forma de *tiempos de seguridad* se insertan entre las franjas de tiempo asignadas. Esto se realiza para reducir la interferencia entre usuarios al permitir la incertidumbre del

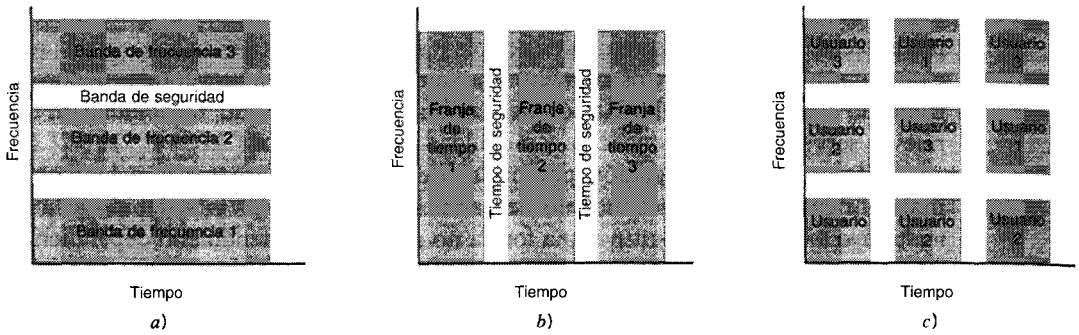


FIGURA 8.1 Ilustración de las ideas detrás de las técnicas de acceso múltiple. *a*) Acceso múltiple por división de frecuencia. *b*) Acceso múltiple por división de tiempo. *c*) Acceso múltiple por salto de frecuencia.

tiempo que se produce por causa de las imperfecciones del sistema, especialmente en esquemas de sincronización.

3. *Acceso múltiple por división de código (CDMA).*
En el FDMA, los recursos del canal se comparten dividiéndolos a lo largo de la coordenada de frecuencias en bandas de frecuencia disjuntas, como se ilustra en la figura 8.1*a*. En el TDMA, la compartición se realiza al dividir los recursos a lo largo de la coordenada del tiempo en franjas de tiempo disjuntas, como se ilustra en la figura 8.1*b*. En la figura 8.1*c* mostramos otra técnica para compartir los recursos del canal al utilizar una combinación híbrida del FDMA y del TDMA, la cual representa una forma específica del acceso múltiple por división de código (CDMA). Por ejemplo, el *salto de frecuencia* puede emplearse para asegurar que durante cada franja de tiempo sucesiva, se reorientan las bandas de frecuencia asignadas a los usuarios de una manera esencialmente al azar. De modo específico, durante la franja de tiempo 1, el usuario 1 ocupa la banda de frecuencia 1, el usuario 2 ocupa la banda de frecuencia 2, el usuario 3, la banda de frecuencia 3, etc. Durante la franja de tiempo 2, el usuario 1 salta a la banda de frecuencia 3, el usuario 2, a la banda de frecuencia 1, el usuario 3, a la banda de frecuencia 2, etc. Un arreglo de estas características tiene la apariencia de los usuarios que están jugando a ganar la silla cuando se interrumpe la música. Una ventaja importante del CDMA con respecto tanto al FDMA como al TDMA es que puede ofrecer una comunicación *segura*. En el tipo de CDMA que se ilustra en la figura 8.1*c*, es posible poner en práctica el mecanismo de salto de frecuencia mediante el uso de una secuencia de pseudo ruido.

4. *Acceso múltiple por división del espacio (SDMA).*
En esta técnica de acceso múltiple, la asignación de recursos se consigue explotando la separación espacial de los usuarios individuales. En particular, se recurre a *antenas de haces múltiples* para separar las señales de radio apuntándolas a lo largo de diferentes direcciones. De este modo se permite que usuarios diferentes tengan acceso simultáneamente al canal en la misma frecuencia o en la misma franja de tiempo.

Estas técnicas de acceso múltiple comparten una característica común: asignan los recursos de comunicación del canal a través del uso de incoherencia (u ortogonalidad en un sentido general) en el tiempo, la frecuencia o el espacio.

Con la disposición de este material básico, estamos listos para explicar algunos sistemas importantes de comunicación multiusuario.

8.3 *Comunicaciones por satélite*

En un sistema de comunicación por satélite geoestacionario[1] se transmite una señal de mensaje desde una estación terrestre por medio de un *enlace ascendente* hasta un satélite, se am-

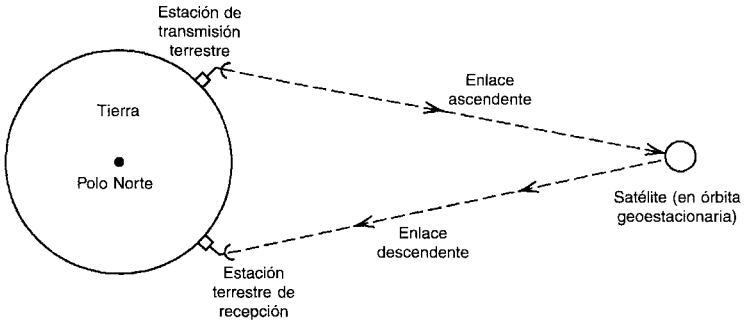


FIGURA 8.2 Sistema de comunicaciones por satélite.

plifica en un *transpondedor* (es decir, circuitería electrónica) a bordo del satélite, y luego se retransmite desde este último por medio de un *enlace descendente* hasta otra estación terrestre, como se ilustra en la figura 8.2. La banda de frecuencia más popular para las comunicaciones por satélite es la de 6 GHz (banda C) para el enlace ascendente y 4 GHz para el descendente. El empleo de esta banda de frecuencia ofrece las siguientes ventajas:

- Equipo de microondas relativamente económico.
- Baja atenuación debido a la lluvia; ésta es la principal causa atmosférica de la degradación de señales.
- Insignificante ruido de fondo del cielo; éste (producto de las emisiones de ruido aleatorias provenientes de las fuentes galáctica, solar y terrestre) alcanza su nivel más bajo entre 1 y 10 GHz.

Sin embargo, la interferencia de radio limita las aplicaciones de satélites de comunicación que operen en la banda de 6/4 GHz, debido a que las frecuencias de transmisión de esta banda coinciden con aquellas utilizadas en los sistemas de microondas terrestres. Este problema se elimina en los más poderosos satélites de comunicaciones de "segunda generación" que operan en la banda de 14/12 GHz (es decir, la banda Ku); además, el uso de estas frecuencias más altas posibilita la construcción de antenas más pequeñas y, por tanto, menos costosas.

El diagrama de bloques de la figura 8.3 muestra los componentes básicos de un canal transpondedor simple de un satélite de comunicación representativo. De manera específica, la salida de la antena de recepción del enlace ascendente se aplica a la conexión en cascada de los siguientes componentes:

- *Filtro pasobanda*, diseñado para separar la señal recibida de entre los diferentes canales de radio.
- *Amplificador de bajo ruido.*

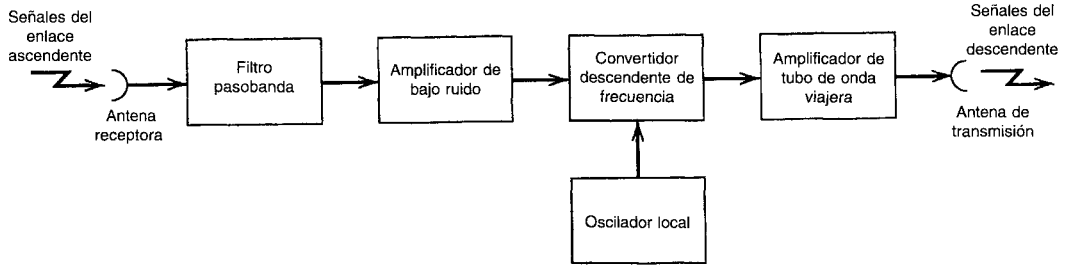


FIGURA 8.3 Diagrama de bloques del transpondedor.

- *Convertidor descendente de frecuencia*, cuyo propósito es convertir la señal de frecuencia de radio (RF) recibida en la frecuencia del enlace descendente que se desea.
- *Amplificador de tubo de onda viajera*, que proporciona alta ganancia sobre una banda ancha de frecuencias. En un tubo de onda viajera (TWT), una señal electromagnética viaja a lo largo de una hélice (es decir, una bobina de alambre en forma de resorte), mientras que los electrones en el haz de alto voltaje se desplazan por la hélice a una velocidad cercana a la de la onda de la señal; el resultado neto es la transferencia de potencia a partir de los electrones a la onda, la cual crece con rapidez cuando la onda de la señal viaja por la hélice.

La configuración del canal que se muestra en la figura 8.3 utiliza una traslación de frecuencia simple. Otras configuraciones de canal realizan la conversión de frecuencia desde el enlace ascendente hasta la frecuencia del enlace descendente en dos etapas: conversión descendente hasta una frecuencia intermedia, seguida de amplificación, y después la conversión ascendente hasta la frecuencia de transmisión deseada.

El retardo del tiempo de propagación se vuelve en particular pronunciado en un canal de satélite debido a las largas distancias implicadas. Específicamente, las señales de voz enviadas por satélite incurren en un retardo de transmisión de casi 270 ms. Por tanto, para señales de voz, cualquier desajuste de impedancia en el extremo receptor de un enlace de satélite produce un *eco* de la voz del hablante, que se escucha de regreso en el extremo transmisor después de un retardo de viaje redondo de aproximadamente 540 ms. Podemos superar este problema utilizando un *cancelador del eco*, que es un dispositivo que sustrae una estimación del eco de la trayectoria de regreso; la eliminación del eco se lleva a cabo por medio de un filtro especial que se adapta por sí solo a las características variables del canal.

El canal de satélite se representa bastante bien por medio de un *modelo de ruido blanco gaussiano aditivo* (AWGN), el cual se aplica tanto a las porciones del enlace ascendente como a las del descendente del sistema de comunicación por satélite. Por tanto, buena parte del material que se presentó en el capítulo 6 acerca de los sistemas pasobanda para la transmisión de datos, con referencia particular a las técnicas de corrimiento de fase por llaveo y corrimiento de frecuencia por llaveo, se aplica de manera directa a las comunicaciones por satélite digitales.

Un transpondedor de satélite difiere de un repetidor de línea visual de microondas convencional en que muchas estaciones terrestres pueden tener acceso al satélite desde localidades ampliamente diferentes sobre la tierra al mismo tiempo (o casi). Esta capacidad es posible utilizando una de las técnicas de acceso múltiple que se explicaron en la sección 8.2. En este contexto podemos hacer las siguientes observaciones:

- En un canal de satélite, la no linealidad del transpondedor es la causa principal de la interferencia entre usuarios. Para enfrentar este serio problema, el amplificador de tubo de onda viajera en el transpondedor se opera a propósito por debajo de su capacidad. En consecuencia, encontramos que en un sistema FDMA la eficiencia de potencia del sistema se reduce en virtud de la *demora de potencia* necesaria del amplificador de tubo de onda viajera.
- En un sistema TDMA, los usuarios tienen acceso al transpondedor de satélite uno a la vez. Por tanto, dicho transpondedor es en estas condiciones capaz de operar muy cerca de la eficiencia de potencia total permitiendo al amplificador de tubo de onda viajera operar hasta la saturación. Esto, a su vez, significa que el TDMA utiliza de manera más eficiente el transpondedor que el FDMA, de lo cual se origina su amplio uso en la puesta en práctica de los sistemas de comunicación digital por satélite.
- El SDMA opera explotando las ubicaciones espaciales de las estaciones terrestres, lo cual se consigue por medio de una *conmutación a bordo*. En forma específica, el transformador está equipado con antenas múltiples, eligiéndose el haz de la antena adecuada para la transmisión de radio hacia una estación terrestre en particular que está demandando el uso del transpondedor.

Además del acceso múltiple, otra capacidad del canal de satélite es la *transmisión* con énfasis en la cobertura de área amplia. Aquí mencionamos los satélites de transmisión, los cuales se caracterizan por su alta transmisión de potencia hasta receptores de bajo costo. Esta

característica se explota en el uso de los *satélites de transmisión directa* (DBS), designados para la recepción casera de servicios de televisión sobre una escala muy amplia. En comparación con las largas estaciones terrenas utilizadas en las comunicaciones por satélite, las estaciones terrestres para los DBS son muy simples y, por tanto, económicas.

8.4 *Análisis del enlace de radio*

Un aspecto importante que se presenta en el diseño de los sistemas de comunicación por satélite es el del análisis del presupuesto del enlace.[2] Como su nombre lo indica, un *presupuesto del enlace*, o más precisamente "presupuesto de la potencia del enlace", es el total de las ganancias y pérdidas en las que se incurre al operar un enlace de comunicación. En particular, la hoja de balance que constituye el presupuesto del enlace proporciona una cuenta detallada de tres elementos ampliamente definidos:

1. Reparto de los recursos disponibles para el transmisor y el receptor.
2. Fuentes responsables de la pérdida de potencia de la señal.
3. Fuentes de ruido.

Reuniendo todos estos elementos en el presupuesto del enlace, finalizamos con un procedimiento de *estimación* para evaluar el desempeño de un enlace de radio, el cual podría ser el enlace ascendente o descendente de un sistema de comunicación por satélite. No obstante, la esencia del análisis del enlace de comunicación que se presenta en esta sección se aplica también a otros enlaces de radio que se fundamentan para su operación en la *línea de visión*. Por esta razón el tratamiento del análisis del enlace de radio que se expone en esta sección es de naturaleza general. La sección concluye con un ejemplo ilustrativo acerca del análisis del presupuesto de un enlace descendente de un sistema de comunicación digital por satélite.

A partir del material que se presentó en el capítulo 6 aprendimos que el desempeño de un sistema de comunicación digital, ante la presencia de ruido del canal modelado como blanco gaussiano aditivo, se define mediante una fórmula que tiene la forma de una curva de "cascada", como se muestra en la figura 8.4. Esta figura muestra la probabilidad del error de símbolo,

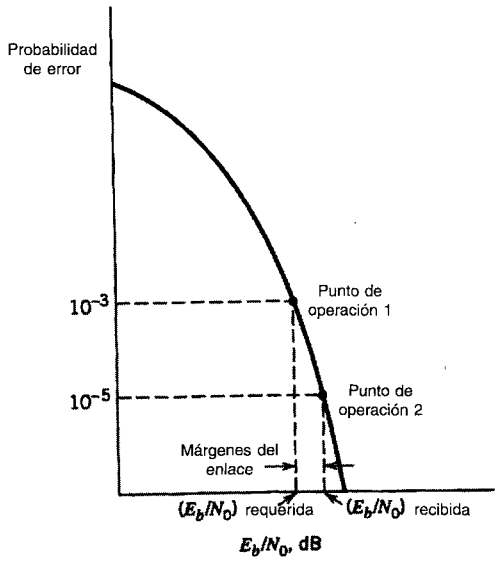


FIGURA 8.4 Curva de "cascada" que relaciona la probabilidad de error con la proporción E_b/N_0.

P_e, graficada en función de la relación de energía de bit a la densidad espectral de ruido, E_b/N_0. Una vez que se ha elegido el esquema de modulación, la primera tarea de diseño consiste en especificar dos valores particulares de E_b/N_0, como se describe a continuación:

1. E_b/N_0 *requerida.*
Suponga, por ejemplo, que la probabilidad preestablecida del error de símbolo es $P_e = 10^{-3}$. Utilizando la curva de cascada de la figura 8.4 relativa al esquema de modulación de interés, se determina la E_b/N_0 requerida para producir la P_e preestablecida. Considere que $(E_b/N_0)_{req}$ denota el valor de la E_b/N_0 obtenida a partir de este cálculo. La P_e preestablecida y la $(E_b/N_0)_{req}$ calculada definen un punto sobre la curva de cascada de la figura 8.4, el cual se designa como el punto de operación 1.

2. E_b/N_0 *recibida.*
Para asegurar la operación *confiable* del enlace de comunicación, el presupuesto de este mismo incluye una medida de seguridad llamada *margen del enlace*. Este último proporciona protección contra cambios e imprevistos. De ese modo el cociente (E_b/N_0) recibido realmente por el sistema es un poco mayor que $(E_b/N_0)_{req}$. Suponga que $(E_b/N_0)_{rec}$ denota la E_b/N_0 real o recibida, lo cual define un segundo punto sobre la curva de la cascada de la figura 8.4, designado como el punto de operación 2. La P_e correspondiente al punto de operación 2 se muestra como 10^{-5} en la figura 8.4 únicamente como ilustración. En cualquier caso, introduciendo un margen del enlace denotado por M, podemos escribir

$$\left(\frac{E_b}{N_0}\right)_{rec} = M\left(\frac{E_b}{N_0}\right)_{req} \tag{8.1}$$

De manera equivalente, al expresar los dos valores de interés de E_b/N_0 en decibeles, es posible definir el margen del enlace como

$$M(dB) = \left(\frac{E_b}{N_0}\right)_{rec}(dB) - \left(\frac{E_b}{N_0}\right)_{req}(dB) \tag{8.2}$$

Naturalmente, cuanto más grande hagamos el margen del enlace M, tanto más confiable resulta el enlace de comunicación. Sin embargo, la confiabilidad creciente del enlace se alcanza a costa de una E_b/N_0 superior.

■ MODELO DE PROPAGACIÓN EN EL ESPACIO LIBRE

El siguiente paso en la formulación del presupuesto del enlace corresponde a calcular la potencia de la señal recibida. De manera natural, este cálculo toma en cuenta todas las ganancias y pérdidas en las que se incurre en la transmisión y recepción de la portadora.

En un sistema de comunicación de radio, la propagación de la señal modulada se consigue por medio de una *antena transmisora*, cuya función es doble:

- Convertir la señal modulada eléctrica en un campo electromagnético. En esta capacidad, la antena transmisora actúa como un transductor "transformador de impedancia" acoplando la impedancia de la antena con la del espacio libre.
- Radiar la energía electromagnética en las direcciones deseadas.

En el receptor tenemos una *antena receptora*, cuya función es la opuesta a la de la antena transmisora: convertir el campo electromagnético en una señal eléctrica a partir de la cual se extrae la señal modulada. Además, la antena receptora se requiera tal vez para suprimir la radiación que proviene de direcciones que no se desean.

Por lo común, el receptor se localiza en el campo lejano de la antena transmisora, en cuyo caso, para todos los propósitos prácticos, es posible que consideremos la antena transmisora como un emisor ficticio sin volumen o *fuente puntual*. Una descripción completa del campo lejano de la fuente puntual requiere del conocimiento del campo electromagnético como una

función del tiempo y el espacio. Sin embargo, en lo que se refiere a los cálculos del enlace, no es necesario un conocimiento tan completo. Más bien, resulta suficiente especificar sólo la variación de la densidad de potencia para la antena.

Por definición, el *vector de Poynting* o *densidad de potencia es la tasa de flujo de energía por área unitaria*; éste tiene las dimensiones de watts por metro cuadrado. El tratamiento de la antena transmisora como una fuente puntual simplifica considerablemente las cosas en el sentido de que la densidad de potencia de una fuente puntual tiene únicamente una componente radial; esto es, la energía radiada surge desde la fuente a lo largo de líneas radiales.

Es útil tener una antena de "referencia" con respecto a la cual sea posible comparar el desempeño de las antenas de transmisión y recepción. La práctica común radica en asumir que la antena de referencia es una *fuente isotrópica*, definida como una *antena omnidireccional (es decir, completamente no direccional) que radia uniformemente en todas direcciones*. Una fuente isotrópica es hipotética, puesto que en realidad todas las antenas de radio tienen cierta directividad, aunque ésta sea pequeña. En consecuencia, la noción de una fuente isotrópica es útil, en especial para fines de comparación de ganancia.

Considere entonces una fuente isotrópica que radia una potencia total denotada por P_t, medida en watts. La potencia radiada pasa uniformemente a través de una esfera de área superficial $4\pi d^2$, donde d es la distancia (en metros) desde la fuente. Por tanto, la densidad en potencia, denotada por $\rho(d)$, está determinada en cualquier punto de la superficie de la esfera por

$$\rho(d) = \frac{P_t}{4\pi d^2} \text{ watts/m}^2 \tag{8.3}$$

La ecuación (8.3) establece que la densidad en potencia varía inversamente con el cuadrado de la distancia desde una fuente puntual. Este enunciado corresponde a la familiar *ley inversa cuadrática* que gobierna la propagación de las ondas electromagnéticas en el espacio libre.

Al multiplicar la densidad de potencia $\rho(d)$ por el cuadrado de la distancia d a la cual ésta se mide, obtenemos una cantidad llamada *intensidad de radiación*, denotada mediante Φ. Es posible escribir en ese caso

$$\Phi = d^2\, \rho(d) \tag{8.4}$$

En tanto que la densidad de potencia $\rho(d)$ se mide en watts por metro cuadrado, la intensidad de la radiación Φ se mide en watts por ángulo sólido unitario (watts por esterradián).

En el caso de una antena de radio de transmisión o recepción característica, la intensidad de la radiación es una función de las coordenadas esféricas θ y ϕ definidas en la figura 8.5. De ese modo, generalmente es posible expresar la intensidad de la radiación como $\Phi(\theta, \phi)$, y hablar

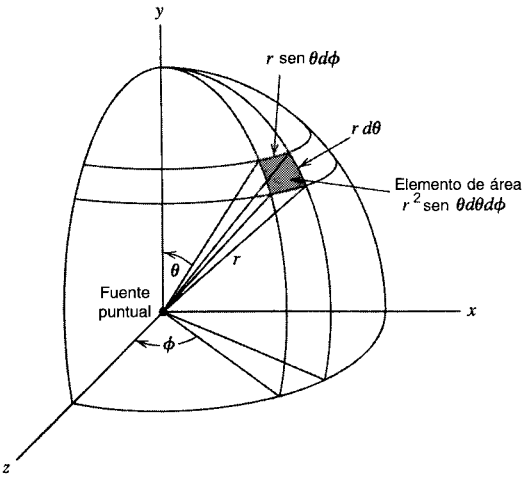


FIGURA 8.5 Ilustración de las coordenadas esféricas de una fuente puntual.

por ello de un *patrón de intensidad de radiación*. La potencia radiada dentro de un ángulo sólido infinitesimal $d\Omega$ está dada por $\Phi(\theta, \phi)\, d\Omega$, donde (refiriéndonos a la figura 8.5)

$$d\Omega = \text{sen}\ \theta\, d\theta\, d\phi \quad \text{esterradián} \tag{8.5}$$

La potencia total radiada es entonces

$$P = \int \Phi(\theta, \phi)\, d\Omega \quad \text{watts} \tag{8.6}$$

que es un enunciado matemático del *teorema de la potencia*. En otras palabras, el teorema de la potencia establece que si se conoce el patrón de intensidad de radiación $\Phi(\theta, \phi)$ para todos los valores del par de ángulos (θ, ϕ), entonces la potencia total radiada está dada por la integral de $\Phi(\theta, \phi)$ sobre un ángulo sólido de 4π esterradianes. La potencia *promedio* radiada por ángulo sólido unitario es

$$P_{\text{pro}} = \frac{1}{4\pi} \int \Phi(\theta, \phi)\, d\Omega$$

$$= \frac{P}{4\pi} \quad \text{watts/esterradián} \tag{8.7}$$

que representa la intensidad de la radiación que produce una fuente isotrópica que radia la misma potencia total P.

Ganancia direccional, direccionalidad y ganancia de potencia[3]

En esas condiciones, la capacidad de una antena para concentrar la potencia radiada en una dirección determinada, como en el caso de la antena de transmisión, o, inversamente, para absorber de modo efectivo la potencia incidente proveniente de esa dirección, como en el caso de la antena receptora, se especifica en términos de su ganancia direccional o direccionalidad. Para una dirección especificada por el par de ángulos (θ, ϕ), la *ganancia direccional* de una antena, denotada por $g(\theta, \phi)$, está definida como *la proporción entre la intensidad de la radiación en esa dirección y la potencia radiada promedio*, según se expresa mediante

$$g(\theta, \phi) = \frac{\Phi(\theta, \phi)}{P_{\text{pro}}}$$

$$= \frac{\Phi(\theta, \phi)}{P/4\pi} \tag{8.8}$$

La *direccionalidad* de una antena, denotada por D, se define como *la proporción entre la intensidad de la radiación máxima proveniente de la antena y la intensidad de la radiación proveniente de una fuente isotrópica*. Esto es, la direccionalidad D es el valor máximo de la ganancia direccional $g(\theta, \phi)$. De ese modo, en tanto que la ganancia direccional de una antena es una función del par de ángulos (θ, ϕ), la direccionalidad D es una constante que se hace máxima en una dirección particular.

La definición de direccionalidad se basa en la forma del patrón de intensidad de radiación $\Phi(\theta, \phi)$; como tal, no incluye el efecto de las imperfecciones de la antena debidas a la pérdida de disipación y al desacoplamiento de impedancia. Una cantidad denominada *ganancia de potencia* implica la eficiencia de radiación de la antena. De manera específica, la ganancia de potencia de una antena, denotada por G, se define como *la proporción entre la intensidad de radiación máxima de la antena y la intensidad de radiación de una fuente isotrópica sin pérdidas, bajo la restricción de que se aplica la misma potencia de entrada en ambas antenas*. Específicamente, al usar $\eta_{\text{radiación}}$ para denotar el *factor de eficiencia de radiación* de la antena, es posible relacionar la ganancia de potencia G con la direccionalidad D como

$$G = \eta_{\text{radiación}} D \tag{8.9}$$

De ese modo, la ganancia de potencia de una antena con respecto a una fuente isotrópica sin pérdida es igual a la direccionalidad si la antena es 100% eficiente (es decir, $\eta_{\text{radiación}} = 1$),

aunque es menor que la direccionalidad si se presenta cualquier pérdida en la antena (es decir, $\eta_{\text{radiación}} < 1$). De aquí en adelante, supondremos que la antena es 100% eficiente y, por tanto, nos referiremos sólo a la ganancia de potencia de la antena.

El concepto de ganancia de potencia, el cual se basa en la forma del patrón de potencia transmitida, puede extenderse a una antena receptora en virtud del *principio de reciprocidad*. Se afirma que una antena será recíproca si el medio de transmisión es lineal, pasivo e isotrópico. Para una estructura de antena determinada, las ganancias de potencia de las antenas de transmisión y recepción son idénticas.

La ganancia de potencia de una antena es el resultado de concentrar la densidad de potencia en una región restringida más pequeña que 4π esterradianes, como se ilustra en la figura 8.6. A la luz de la imagen que se ilustra en esta figura, es posible introducir los siguientes dos parámetros:

1. *Potencia radiada efectiva referida a una fuente isotrópica* (*EIRP*); dicha potencia se define como el *producto de la potencia transmitida, P_t, y la ganancia de potencia de la antena transmisora, G_t,* como indica
$$\text{EIRP} = P_t G_t \quad \text{watts} \tag{8.10}$$
2. *Ancho del haz de la antena*, que representa una medida "planar" del ángulo sólido de visión de la antena; el ancho del haz, en grados o radianes, se define como *el ángulo que subtienden los dos puntos sobre el lóbulo principal del patrón de potencia del campo al cual la potencia pico del mismo se reduce en 3 dB*. Cuanto más alta sea la ganancia de potencia de la antena, tanto más estrecho resulta el ancho del haz de la misma.

Otro asunto de interés que se distingue a partir de la figura 8.6 es el referente a los *lóbulos laterales* del patrón de la potencia del campo. Desafortunadamente, toda antena física tiene lóbulos laterales, los cuales son responsables de absorber las radiaciones de interferencia indeseables.

Apertura efectiva

Un término que tiene un significado especial en una antena de recepción es la *apertura efectiva* de la antena, que se define como *la proporción entre la potencia disponible en las terminales de la antena y la potencia por área unitaria de la onda electromagnética incidente polarizada en forma apropiada*. La apertura efectiva, denotada por medio de A, se define en términos de la ganancia de potencia de la antena G como

$$A = \frac{\lambda^2}{4\pi} G \tag{8.11}$$

donde λ es la *longitud de onda* de la portadora. La longitud de onda λ y la frecuencia f se relacionan en forma recíproca como

$$\lambda = \frac{c}{f} \tag{8.12}$$

donde c es la velocidad de la luz (aproximadamente igual a 3×10^8 m/s).

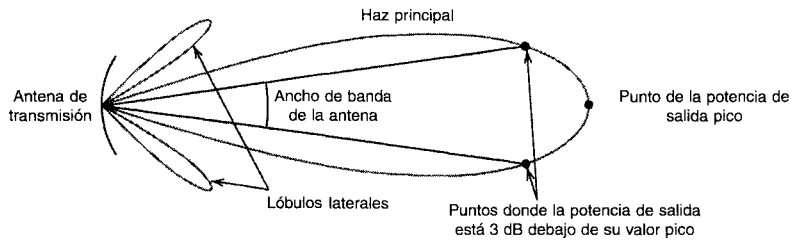


FIGURA 8.6 Ilustración de la concentración de la densidad de potencia de una antena transmisora dentro de una región más pequeña que 4π esterradianes.

El término apertura efectiva tiene un significado particular en el contexto de las antenas reflectoras y las antenas electromagnéticas que se caracterizan por una apertura perfectamente definida. En estas antenas, *la proporción entre la apertura efectiva de la antena y su apertura física es una medida directa de la eficiencia de la apertura de la antena*, η_{apertura}, en la potencia radiada hacia una dirección que se desea o la potencia absorbida a partir de esa dirección. Los valores nominales para la eficiencia η_{apertura} de las antenas de reflector se encuentran en la gama de 45 a 75 por ciento.

Ecuación de Friis del espacio libre

Teniendo a nuestra disposición este material introductorio acerca de las antenas, estamos listos ahora para formular la ecuación de propagación básica para un enlace de comunicación de radio. Considere una antena transmisora con una EIRP definida en la ecuación (8.10). Recurriendo a la ley inversa cuadrática de la ecuación (8.3), es posible expresar la densidad de potencia de la antena de transmisión como $\text{EIRP}/4\pi d^2$, donde d es la distancia entre las antenas receptora y transmisora. La potencia P_r que absorbe la antena de recepción es el producto de esta densidad de potencia y el área efectiva de la antena denotada por A_r:

$$\begin{aligned} P_r &= \left(\frac{\text{EIRP}}{4\pi d^2}\right) A_r \\ &= \frac{P_t G_t A_r}{4\pi d^2} \quad \text{watts} \end{aligned} \tag{8.13}$$

De acuerdo con el principio de reciprocidad, podemos utilizar la ecuación (8.11) para expresar el área efectiva de la antena receptora como

$$A_r = \frac{\lambda^2}{4\pi} G_r$$

donde G_r es la ganancia de potencia de la antena receptora. Sustituyendo esta fórmula para A_r en la ecuación (8.13), se expresa la potencia de la señal recibida en la forma equivalente

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \tag{8.14}$$

Esta última expresión recibe el nombre de *ecuación de Friis del espacio libre*.[4]

La *pérdida de trayectoria*, PT, que representa la "atenuación" de la señal en decibeles a lo largo de todo el enlace de comunicación, se define como *la diferencia (en decibeles) entre la potencia de la señal transmitida* P_t *y la potencia de la señal recibida* P_r, según se expresa por medio de

$$\begin{aligned} \text{PT} &= 10 \log_{10}\left(\frac{P_t}{P_r}\right) \\ &= -10 \log_{10}(G_t G_r) + 10 \log_{10}\left(\frac{4\pi d}{\lambda}\right)^2 \end{aligned} \tag{8.15}$$

El signo menos asociado con el primer término en la ecuación (8.15) implica el hecho de que este término representa una "ganancia". El segundo término, debido a la colección de términos $(4\pi d/\lambda)^2$, se conoce como *pérdida del espacio libre*, denotado por $L_{\text{espacio libre}}$. Advierta que al aumentar la distancia d que separa a las antenas de recepción y transmisión se produce un incremento en la pérdida del espacio libre, el cual, a su vez, nos obliga a operar el enlace de comunicación de radio a frecuencias inferiores a fin de mantener la pérdida de trayectoria en un nivel manejable.

La ecuación de Friis del espacio libre nos permite calcular la pérdida de trayectoria para valores especificados de ganancias de potencia G_t y G_r, la longitud de onda de la portadora λ y

la distancia d. Para completar el análisis del presupuesto del enlace, necesitamos calcular la potencia promedio del ruido en la señal recibida, lo cual se considera a continuación.

Figura de ruido

Para efectuar un análisis del ruido en el receptor de un sistema de comunicación, necesitamos una medida conveniente del desempeño ante el ruido de un dispositivo lineal de dos puertos. Una medida de este tipo la proporciona la llamada *figura de ruido*. Considere un dispositivo lineal de dos puertos conectado a una fuente de señales de impedancia interna $Z(f) = R(f) + jX(f)$ a la entrada, como en la figura 8.7. El voltaje del ruido $v(t)$ representa el ruido térmico asociado con la resistencia interna $R(f)$ de la fuente. El ruido de salida del dispositivo está integrado por dos contribuciones, una debida a la fuente y la otra producto del propio dispositivo. Definimos *la potencia disponible del ruido de salida en una banda de ancho Δf centrada a la frecuencia f como la potencia promedio máxima del ruido en esta banda, que se obtiene a la salida del dispositivo*. La potencia de ruido máxima que el dispositivo de dos puertos puede entregar a una carga externa se obtiene cuando la impedancia de carga es la conjugada compleja de la impedancia de salida del dispositivo, esto es, cuando la resistencia está acoplada y la reactancia desintonizada. Definimos la *figura de ruido del dispositivo de dos puertos como la proporción entre la potencia total disponible del ruido de salida (debida al dispositivo y a la fuente) por ancho de banda unitario y la porción de ella debida exclusivamente a la fuente.*

Dejemos que la densidad espectral de la potencia total de ruido disponible de la salida del dispositivo sea $S_{NO}(f)$, y que la densidad espectral de la potencia de ruido disponible debida a la fuente a la entrada del dispositivo corresponda a $S_{NS}(f)$. Dejemos además que $G(f)$ denote la *ganancia de potencia disponible* del dispositivo de dos puertos, definida como la *proporción entre la potencia de la señal disponible a la salida del dispositivo y la potencia de la señal disponible de la fuente cuando la señal es una onda senoidal de frecuencia f*. En ese caso es posible expresar la figura de ruido F del dispositivo como

$$F = \frac{S_{NO}(f)}{G(f)S_{NS}(f)} \tag{8.16}$$

Si el dispositivo estuviera exento de ruido, $S_{NO}(f) = G(f)S_{NS}(f)$, y la figura de ruido sería entonces la unidad. Sin embargo, en un dispositivo físico, $S_{NO}(f)$ es mayor que $G(f)S_{NS}(f)$, de modo que la figura de ruido siempre es mayor que la unidad. Esta figura se expresa comúnmente en decibles, esto es, como $10 \log_{10} F$.

La figura de ruido puede expresarse también en una forma alternativa. Sea $P_S(f)$ la potencia disponible de la señal proveniente de la fuente, la cual es la potencia de señal máxima promedio que es posible obtener. En el caso de una fuente que proporciona una componente de

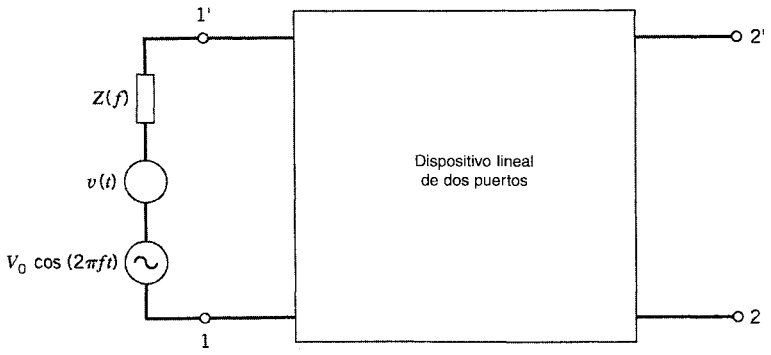


Figura 8.7 Dispositivo lineal de dos puertos.

señal de una sola frecuencia con un voltaje en circuito abierto $V_0 \cos(2\pi f t)$, la potencia disponible de la señal se obtiene cuando la carga conectada a la fuente es

$$Z^*(f) = R(f) - jX(f)$$

donde el asterisco denota conjugación compleja. Bajo esta condición, encontramos que

$$\begin{aligned} P_S(f) &= \left[\frac{V_0}{2R(f)}\right]^2 R(f) \\ &= \frac{V_0^2}{4R(f)} \end{aligned} \tag{8.17}$$

La potencia de la señal disponible a la salida del dispositivo es entonces

$$P_O(f) = G(f)P_S(f) \tag{8.18}$$

En ese caso, al multiplicar tanto el numerador como el denominador del segundo miembro de la ecuación (8.16) por $P_S(f)\ \Delta(f)$, obtenemos

$$\begin{aligned} F &= \frac{P_S(f)S_{NO}(f)\,\Delta f}{G(f)P_S(f)S_{NS}(f)\,\Delta f} \\ &= \frac{P_S(f)S_{NO}(f)\,\Delta f}{P_O(f)S_{NS}(f)\,\Delta f} \\ &= \frac{\rho_S(f)}{\rho_O(F)} \end{aligned} \tag{8.19}$$

donde

$$\rho_S(f) = \frac{P_S(f)}{S_{NS}(f)\,\Delta f} \tag{8.20}$$

$$\rho_O(f) = \frac{P_O(f)}{S_{NO}(f)\,\Delta f} \tag{8.21}$$

Nos referimos a $\rho_S(f)$ como la *relación de la señal disponible a ruido de la fuente* y a $\rho_O(f)$ como la *relación de la señal disponible a ruido a la salida del dispositivo*, medidas ambas en una banda estrecha de ancho Δf centrado en f. Puesto que la figura de ruido siempre es mayor que la unidad, se sigue a partir de la ecuación (8.19) que la relación señal a ruido siempre disminuye con la amplificación, lo cual constituye un resultado importante.

La figura de ruido F es una función de la frecuencia de operación f; por tanto, se le conoce como la *figura de ruido de punto*. En contraste, es posible definir una *figura de ruido promedio* F_0 de un dispositivo de dos puertos como la proporción entre la potencia de ruido total a la salida del dispositivo y la potencia de ruido de salida debida exclusivamente a la fuente. Esto es,

$$F_0 = \frac{\int_{-\infty}^{\infty} S_{NO}(f)\,df}{\int_{-\infty}^{\infty} G(f)S_{NS}(f)\,df} \tag{8.22}$$

Es claro que en el caso del ruido térmico en el circuito de entrada con $R(f)$ constante y ganancia constante a lo largo de una banda fija con ganancia cero a otras frecuencias, la figura de ruido de punto F y la figura de ruido promedio F_0 son idénticas.

Temperatura de ruido equivalente

Una desventaja de la figura de ruido F es que cuando se usa para comparar dispositivos de ruido bajo, los valores que se obtienen son en su totalidad cercanos a la unidad, lo que hace que la

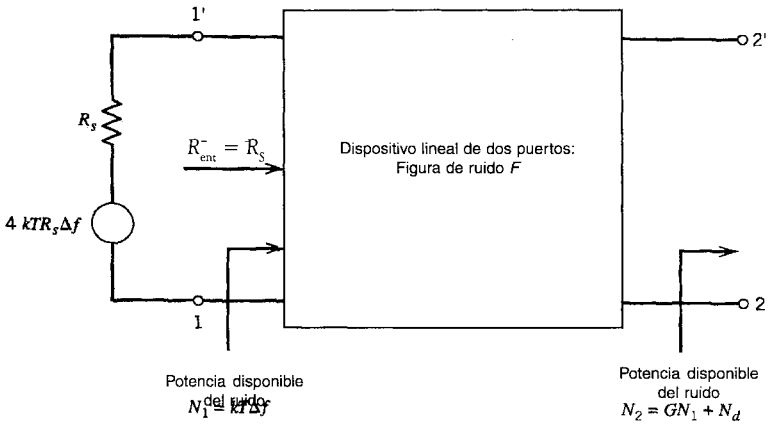


FIGURA 8.8 Dispositivo lineal de dos puertos acoplado a la resistencia interna de una fuente conectada a la entrada.

comparación sea bastante difícil. En su caso, es preferible utilizar la *temperatura de ruido equivalente*. Considere un dispositivo lineal de dos puertos cuya resistencia de entrada se acopla a la resistencia interna de la fuente, como se ilustra en la figura 8.8. En este diagrama hemos incluido también el generador del voltaje de ruido asociado con la resistencia interna R_s de la fuente. El valor cuadrático medio de este voltaje de ruido corresponde a $4kTR_s\,\Delta f$, donde k es la *constante de Boltzmann*. Por tanto, la potencia de ruido disponible a la entrada del dispositivo es

$$N_1 = kT\,\Delta f \tag{8.23}$$

Sea N_d la potencia de ruido que aporta el dispositivo de dos puertos a la potencia de ruido de salida disponible total N_2. Definimos N_d como

$$N_d = GkT_e\,\Delta f \tag{8.24}$$

donde G es la ganancia de potencia disponible del dispositivo y T_e es su temperatura de ruido equivalente. Se sigue entonces que la potencia total del ruido de salida es

$$\begin{aligned} N_2 &= GN_1 + N_d \\ &= Gk(T + T_e)\,\Delta f \end{aligned} \tag{8.25}$$

La figura de ruido del dispositivo es entonces (vea el puerto de salida de la figura 8.8)

$$F = \frac{N_2}{N_2 - N_d} = \frac{T + T_e}{T} \tag{8.26}$$

Al resolver para la temperatura del ruido equivalente:

$$T_e = T(F - 1) \tag{8.27}$$

La figura de ruido F se mide en condiciones de entrada acopladas y con la fuente de ruido a temperatura T. Por convención la temperatura T se toma como la "ambiente", esto es, 290 K, donde K representa "grado Kelvin".

Conexión en cascada de redes de dos puertos

Muchas veces resulta necesario evaluar la figura de ruido de una conexión en cascada de redes de dos puertos cuyas figuras de ruido individuales se conocen. Considere la figura 8.9,

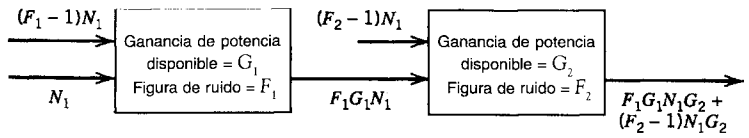


FIGURA 8.9 Una cascada de dos redes de dos puertos con ruido.

consistente en un par de redes de dos puertos de figuras de ruido F_1 y F_2 y ganancias de potencia G_1 y G_2, conectadas en cascada. Se supone que los dispositivos están acoplados y que la figura de ruido F_2 de la segunda red se define suponiendo una potencia de ruido de entrada N_1.

A la entrada de la primera red tenemos una potencia de ruido N_1 que aporta la fuente, más una potencia de ruido equivalente $(F_1 - 1)N_1$ que aporta la propia red. La potencia del ruido de salida proveniente de la primera red es entonces $F_1N_1G_1$. Agregada a esta potencia de ruido, a la entrada de la segunda red tenemos una potencia adicional equivalente $(F_2 - 1)N_1$ aportada por la misma segunda red. La potencia de salida del ruido a partir de esta segunda red es entonces igual a $F_1G_1N_1G_2 + (F_2 - 1)N_1G_2$. Es posible considerar la figura de ruido F como la proporción entre la potencia real de salida del ruido y la potencia de salida del ruido suponiendo las redes sin ruido. Por tanto, podemos expresar la figura de ruido total de la conexión en cascada de la figura 8.9 como

$$F = \frac{F_1G_1N_1G_2 + (F_2 - 1)N_1G_2}{N_1G_1G_2}$$
$$= F_1 + \frac{F_2 - 1}{G_1} \tag{8.28}$$

El resultado se extiende sin dificultad a la conexión en cascada de cualquier número de redes de dos puertos, en la forma indicada por

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2} + \frac{F_4 - 1}{G_1G_2G_3} + \cdots \tag{8.29}$$

donde F_1, F_2, F_3, ..., son las figuras de ruido individuales y G_1, G_2, G_3, ..., son las ganancias de potencia disponible, respectivamente. La ecuación (8.29) muestra que si la primera etapa de la conexión en cascada en la figura 8.9 tiene una ganancia elevada, la figura total de ruido F está dominada por la figura de ruido de la primera etapa.

En forma correspondiente, podemos expresar la temperatura del ruido equivalente total de la conexión en cascada de cualquier número de redes de dos puertos con ruido de la manera siguiente:

$$T_e = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1G_2} + \frac{T_4}{G_1G_2G_3} + \cdots \tag{8.30}$$

donde T_1, T_2, T_3, ..., son las temperaturas de ruido equivalente de las redes individuales, y G_1, G_2, G_3, ..., son las ganancias de potencia, respectivamente. La ecuación (8.30) se conoce como la *fórmula de Friis*. Aquí de nuevo advertimos que si la ganancia G_1 de la primera etapa es elevada, la temperatura del ruido equivalente T_e está dominada por la de la primera etapa.

▶ EJEMPLO 8.1 Temperatura del ruido del receptor de la terminal terrestre

La figura 8.10 presenta un receptor de terminal terrestre característico, compuesto por un amplificador de radio frecuencia (RF) de bajo ruido (LNA), un convertidor descendente de frecuencia

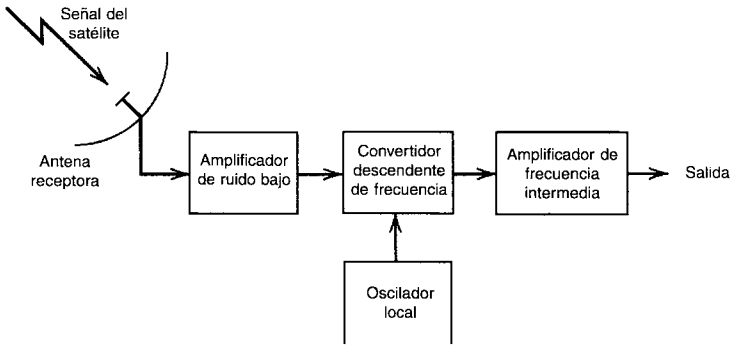


FIGURA 8.10 Diagrama de bloques del receptor de la terminal terrestre.

(mezclador) y un amplificador de frecuencia intermedia (FI). Las temperaturas de ruido equivalente de estos componentes, incluyendo la antena de recepción, son

$$T_{\text{antena}} = 50\ K$$
$$T_{\text{RF}} = 50\ K$$
$$T_{\text{mezclador}} = 500\ K$$
$$T_{FI} = 1000\ K$$

Las ganancias de potencia disponibles de los dos amplificadores son

$$G_{\text{RF}} = 200 = 23\ \text{dB}$$
$$G_{\text{FI}} = 1000 = 30\ \text{dB}$$

Para calcular la temperatura del ruido equivalente del receptor utilizamos la ecuación (8.30), obteniendo

$$T_e = T_{\text{antena}} + T_{\text{RF}} + \frac{T_{\text{mezclador}} + T_{\text{FI}}}{G_{\text{RF}}}$$
$$= 50 + 50 + \frac{500 + 1000}{200}$$
$$= 107.5\ \text{K}$$

EJEMPLO 8.2 Análisis del presupuesto del enlace descendente de un sistema de comunicación digital por satélite

En un sistema de comunicación digital por satélite, uno de los elementos principales en el diseño y análisis totales del sistema es el presupuesto de la potencia del enlace descendente, el cual suele ser más crítico que el presupuesto de potencia del enlace ascendente debido a las restricciones prácticas impuestas en la potencia del primero y en el tamaño de la antena de satélite. El ejemplo que se presenta aquí considera un análisis del presupuesto del enlace descendente de muestra, suponiendo que cualquier potencia del enlace ascendente que se requiera (dentro de límites) se encuentra disponible para la operación satisfactoria del sistema.

El parámetro crítico por calcular es la *proporción de la potencia de la portadora recibida y la densidad espectral de ruido*, denotada por C/N_0. De acuerdo con la ecuación de Friis del espacio libre (8.14), la potencia promedio recibida en la terminal terrestre con respecto a la potencia promedio P_t transmitida por el satélite es

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2$$

donde, en este ejemplo, G_t es la ganancia de potencia de la antena de satélite, G_r es la ganancia de potencia de la antena receptora de la terminal terrestre, λ es la longitud de onda de la portadora correspondiente al enlace descendente, y d es la distancia entre el satélite y la terminal terrena. Dado que la temperatura de ruido equivalente del sistema es T_e, es posible utilizar la ecuación (1.94) del capítulo 1 para expresar la densidad espectral del ruido N_0 como kT_e, donde k es la constante de Boltzmann. Además, de la ecuación (8.10) advertimos que P_tG_t es igual a la EIRP del satélite. Por tanto, al dividir P_r entre N_0, podemos expresar la proporción C/N_0 para el enlace descendente como

$$\left(\frac{C}{N_0}\right)_{\text{enlace descendente}} = (\text{EIRP})_{\text{satélite}}\left(\frac{G_r}{T_e}\right)_{\text{terminal terrestre}}\left(\frac{\lambda}{4\pi d}\right)^2\frac{1}{k} \tag{8.31}$$

Para un sistema de satélite determinado, la pérdida en el espacio libre $(4\pi d/\lambda)^2$ es una constante. Viendo el sistema desde la terminal terrestre, observamos, de acuerdo con la ecuación (8.31), que la proporción (C/N_0) es proporcional a G_r/T_e; por tanto, es posible utilizar el cociente G_r/T_e para valorar la "calidad" de la terminal terrestre; éste suele abreviarse como la *proporción* G/T, la cual se conoce como la *figura de mérito* de la terminal terrestre receptora. De tal modo, al reescribir la fórmula (8.31) para la proporción (C/N_0) medida en decibeles, podemos expresarla como la suma de las ganancias y pérdidas como se particularizan a continuación:

1. $(\text{EIRP})_{\text{satélite}}$, medida en dBW, donde dBW denota decibeles con referencia a 1 watt, esto es, 0 dBW.
2. $(G/T)_{\text{terminal terrestre}}$, medida en dB/K, donde K se refiere a grado Kelvin.
3. $L_{\text{espacio libre}}$, denotando la pérdida del espacio libre $10\log_{10}(4\pi d/\lambda)^2$ en dB.
4. $-10\log_{10} k$, que representa la ganancia en dBW/K-Hz debida a la división entre la constante de Boltzmann $k = 1.38 \times 10^{-23}$ joule/K.

La tabla 8.1 presenta los valores de estos cuatro términos para el enlace descendente de un sistema de comunicación digital por satélite local y representativo, basado en lo siguiente:

1. El transpondedor se opera a su máxima potencia de salida (es decir, no se emplea demora de potencia), produciendo una EIRP de 46.5 dBW.
2. La terminal terrestre receptora utiliza una antena de disco de 2m con una ganancia de potencia G = 45 dB, y el receptor se configura como en el ejemplo 8.1 con temperatura equivalente $T = 107.5$ K. Por tanto,

$$\begin{aligned}\frac{G}{T} &= 45 - 10\log_{10} 107.5\\ &= 45 - 20.3\\ &= 24.7\ dB/K\end{aligned}$$

3. La pérdida en el espacio libre es

$$L_{\text{espacio libre}} = 92.4 + 20\log_{10} f + 20\log_{10} d \text{ dB} \tag{8.32}$$

TABLA 8.1 ***Presupuesto de potencia del enlace descendente para el ejemplo 8.2***

Variable	*Valor*
EIRP	+46.5 dBW
Proporción G/T	+24.7 dB/K
Pérdida del espacio libre	−206 dB
Constante de Boltzmann	+228.6 dBW/K-Hz
C/N_0	93.8 dB-Hz

donde la frecuencia de la portadora del enlace descendente f está en GHz y la distancia d entre el satélite y la terminal terrestre se mide en kilómetros. En el caso de un satélite geoestacionario, la distancia entre el satélite y una terminal terrestre se encuentra en el intervalo de 36,000 a 41,000 km. De ese modo, eligiendo d = 40,000 km y suponiendo f = 12 GHz, el uso de la ecuación (8.32) produce

$$\begin{aligned} L_{\text{espacio libre}} &= 92.4 + 20 \log_{10} 12 + 20 \log_{10} 40{,}000 \\ &= 92.4 + 21.6 + 92.0 \\ &= 206 \text{ dB} \end{aligned}$$

4. Con la constante de Boltzmann $k = 1.39 \times 10^{-23}$ joule/K, su contribución a la proporción C/N_0 es

$$\begin{aligned} -10 \log_{10} k &= 10 \log_{10} 1.38 \times 10^{-23} \\ &= 228.6 \text{ dBW/K - Hz} \end{aligned}$$

Totalizando las ganancias y pérdidas, obtenemos entonces

$$\left(\frac{C}{N_0}\right)_{\text{enlace descendente}} = 93.8 \text{ dB - Hz}$$

El valor del enlace descendente "recibido" de la proporción (C/N_0) puede expresarse también en términos del valor "requerido" de la relación energía de bit a densidad espectral de ruido, $(E_b/N_0)_{\text{req}}$ dB, en la terminal terrestre receptora como [vea la ecuación (8.2)]

$$\left(\frac{C}{N_0}\right)_{\text{enlace descendente}} = \left(\frac{E_b}{N_0}\right)_{\text{req}} + 10 \log_{10} M + 10 \log_{10} R \text{ dB} \tag{8.33}$$

donde $10 \log_{10} M$ es el margen del enlace en decibeles, y R es la velocidad de datos en b/s. El margen del enlace permite excesos en las pérdidas por lluvia en la propagación y en otras degradaciones de potencia. Por lo común, el margen del enlace se elige como 4 dB para la banda C, de 6 dB para la banda *Ku* y más alto para las frecuencias superiores de la banda *K* debido a las mayores pérdidas por lluvia. Para la operación de la frecuencia de la banda *Ku* de 12 GHz, elegimos un margen del enlace de 6 dB. Por tanto, al utilizar el valor C/N_0 = 93.8 dB-Hz calculado para el presupuesto del enlace, el margen de este mismo correspondiente a $10 \log_{10} M$ = 6 dB, y suponiendo que $(E_b/N_0)_{\text{req}}$ = 12.5 dB, el uso de la ecuación (8.33) produce

$$\begin{aligned} 10 \log_{10} R &= 93.8 - 12.5 - 6 \\ &= 75.3 \end{aligned}$$

Por tanto,

$$R = 33.9 \text{ Mb/s}$$

Suponiendo el uso de PSK-8 coherente para la transmisión de datos digitales por el satélite, y sustituyendo (E_b/N_0) = 12.5 dB en la ecuación (6.47) del capítulo 6, encontramos que la probabilidad del error de símbolo $P_e = 0.6 \times 10^{-3}$.

En resumen, el sistema de comunicación digital por satélite analizado en este ejemplo permite, en las peores condiciones de operación, la transmisión de datos sobre el enlace descendente a una velocidad R = 33.9 Mb/s y con una probabilidad del error de símbolo $P_e = 0.6 \times 10^{-3}$, suponiendo el uso de PSK de fase 8. ◀

8.5 *Comunicaciones inalámbricas*

En esta sección estudiamos el segundo tipo de sistema de comunicación de radio multiusuario, a saber, las *comunicaciones inalámbricas*, las cuales son un sinónimo del *radio móvil*. Este último término suele significar formas que incluyen comunicación inalámbrica en interiores y exterio-

res donde un transmisor o receptor de radio es capaz de moverse, sin que importe si en realidad se mueve o no. Debido a la naturaleza estocástica del canal de radio móvil, su caracterización exige el uso de mediciones prácticas y de análisis estadístico. El objetivo de una evaluación de este tipo es cuantificar dos factores de interés principal:

1. *Intensidad media de la señal*, lo cual nos permite predecir la mínima potencia necesaria para irradiar desde el transmisor de modo que se proporcione una calidad aceptable de la cobertura sobre un área de servicio predeterminada.
2. *Variabilidad de la señal*, lo cual caracteriza la naturaleza de desvanecimiento del canal.

Nuestro interés específico en las comunicaciones inalámbricas se encuentra en el contexto del *radio celular*[5] que cuenta con la capacidad inherente de la movilidad de construcciones en la red telefónica. Con dicha capacidad, un usuario puede moverse con libertad dentro de un área de servicio y comunicarse simultáneamente con cualquier suscriptor telefónico en el mundo. Un modelo idealizado del sistema de radio celular, ilustrado en la figura 8.11, consiste en un arreglo de *celdas* hexagonales con una *estación base* localizada en el centro de cada celda; una celda característica tiene un radio de 1 a 12 millas. La función de las estaciones de base es actuar como una interfase entre los *suscriptores móviles* y el sistema de radio celular. Las estaciones de la base se conectan a sí mismas a un *centro de conmutación* mediante alambrados dedicados.

El centro de conmutación móvil tiene dos funciones importantes. Primero, actúa como interfase entre el sistema de radio celular y la red pública de conmutación telefónica. Segundo, lleva a cabo la supervisión y el control completos de las comunicaciones móviles. Realiza la última función supervisando la relación señal a ruido de una llamada en progreso, de acuerdo a como se mide en la estación base en comunicación con el suscriptor móvil implicado en la llamada. Cuando la SNR disminuye por debajo de un umbral preestablecido, lo cual ocurre cuando el suscriptor móvil abandona su celda o cuando se desvanece el canal de radio, éste se conmuta a otra estación de base. Este proceso de conmutación, denominado *escalonamiento* o *desatención*, está enviado para mover un suscriptor móvil de una estación base a otra durante una llamada de un modo transparente, es decir, sin interrupción del servicio.

El concepto celular se fundamenta en dos rasgos esenciales, que se describen en seguida:

1. *Reuso de frecuencia.* El término *reuso de frecuencia* se refiere al empleo de canales de radio sobre la misma frecuencia de portadora para cubrir diferentes áreas, las cuales están sepa-

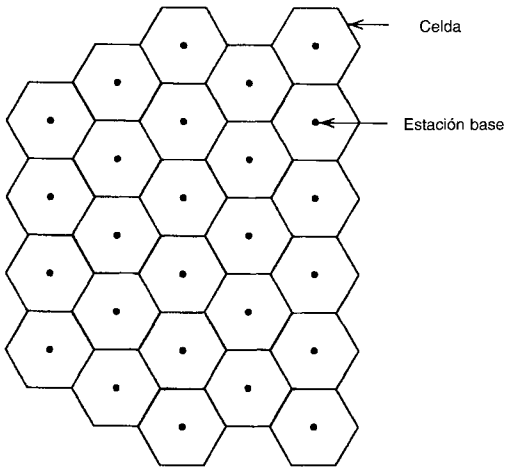


FIGURA 8.11 Modelo idealizado del radio celular.

radas físicamente entre sí lo suficiente para asegurar que la *interferencia de cocanal* no es objetable. De ese modo, en lugar de cubrir un área local completa a partir de un solo transmisor con alta potencia a una altura considerable, el reuso de frecuencia posibilita conseguir dos objetivos de sentido común: mantener la potencia transmitida de cada estación base en un mínimo y ubicar las antenas de las estaciones base justo lo suficientemente elevadas para proporcionar la cobertura del área de las celdas respectivas.

2. *División de celdas*. Cuando la demanda del servicio supera el número de canales asignados a una celda particular, la división de celdas se utiliza para manejar el crecimiento adicional en el tráfico dentro de la celda particular. De modo específico, la división de celda implica una revisión de sus fronteras, de manera que el área local considerada formalmente como una celda única, pueda ahora contener varias celdas más pequeñas y usar los complementos de canal de las nuevas. Estas celdas, que tienen un radio más pequeño que las originales, se denominan *microceldas*. La potencia del transmisor y la altura de la antena de las nuevas estaciones base se reducen en forma correspondiente, y el mismo conjunto de frecuencia se reutiliza de acuerdo con un nuevo plan.

En un modelo hexagonal de un sistema de radio celular, es posible explotar las propiedades básicas de la geometría celular hexagonal para desplegar un plan de asignación del canal de radio que determine qué conjunto de canales debe asignarse a qué celda. Empezamos con dos enteros i y $j (i \geq j)$, denominados *parámetros de corrimiento*, los cuales se predeterminan de alguna manera. Advertimos que con una geometría celular hexagonal hay seis "cadenas" de hexágonos que emanan de cada hexágono y que se extienden en diferentes direcciones. Por tanto, partiendo de cualquier celda como referencia, encontramos las *celdas de cocanal* más cercanas procediendo del siguiente modo:

- Movemos i celdas a lo largo de cualquier cadena de hexágonos, girando 60 grados en sentido contrario al de las manecillas del reloj, y movemos j celdas a lo largo de la cadena que se ubica en esta nueva dirección. Las j celdas localizadas de esta manera y las celdas de referencia constituyen el conjunto de celdas de cocanal.

Este procedimiento se repite para una celda de referencia diferente, hasta que se cubren todas las celdas en el sistema. La figura 8.12 ilustra la aplicación de este procedimiento para una celda de referencia simple y el ejemplo de $i = 2$ y $j = 2$.

En Estados Unidos la banda de frecuencias de radio que se asigna al sistema celular es la de 800–900 MHz. La sub-banda de 824–849 MHz se utiliza para recibir señales de las unidades móviles, y la sub-banda de 869–894 MHz se emplea para transmitir señales a las unidades

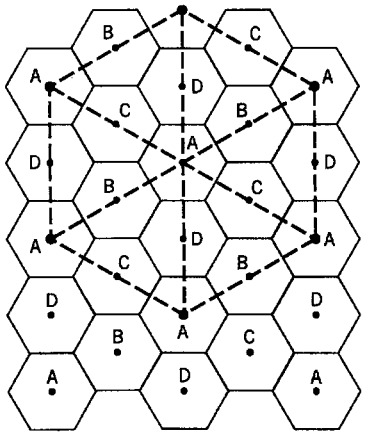


FIGURA 8.12 Ilustración de la determinación de las celdas de cocanal.

móviles. La utilización de estas frecuencias relativamente altas tiene el rasgo benéfico de proporcionar una buena cobertura portátil al penetrar las construcciones. En Europa y otros lados la banda de la base móvil y las sub-bandas de la base móvil se invierten.

■ EFECTOS DE PROPAGACIÓN[6]

Los principales problemas de propagación que se encuentran en el uso del radio celular en áreas con construcciones se deben a que la antena de la unidad móvil puede ubicarse bastante abajo de los edificios circundantes. Planteándolo en términos simples, no hay trayectoria "de línea de visión" hacia la estación base. En cambio, la propagación de radio ocurre principalmente a partir de la dispersión en las superficies de las construcciones de los alrededores y mediante la difracción sobre y/o alrededor de ellas, como se ilustra en la figura 8.13. El punto importante por advertir en la figura 8.13 es que la energía llega a la antena de recepción por conducto de más de una trayectoria. Por tanto, hablamos de un *fenómeno de trayectoria múltiple* en el que diversas ondas de radio entrantes llegan a su destino provenientes de diferentes direcciones y con diferentes retardos de tiempo.

Para entender la naturaleza del fenómeno de multitrayectoria, considere primero un ambiente de multitrayectoria "estático" que incluye un receptor estacionario y una señal transmitida que consiste en una señal de banda angosta (por ejemplo, portadora senoidal no modulada). Supóngase que al receptor arriban en forma secuencial dos versiones atenuadas de la señal transmitida. El efecto del retardo de tiempo diferencial es introducir un corrimiento de fase relativo entre las dos componentes de la señal recibida. Es posible entonces identificar uno de los dos casos extremos que pueden surgir:

- El corrimiento de fase relativo es cero, en cuyo caso las dos componentes se suman constructivamente, como se ilustra en la figura 8.14*a*.
- El corrimiento de fase relativo es de 180 grados, en cuyo caso las dos componentes se suman destructivamente, como se ilustra en la figura 8.14*b*.

También podemos recurrir a los *fasores* para demostrar los efectos constructivos y destructivos de la multitrayectoria, como se ilustra, respectivamente, en las figuras 8.15*a* y 8.15*b*. Advierta que en el ambiente de multitrayectoria estático que se describe aquí, la amplitud de la señal recibida no varía con el tiempo.

Considere a continuación un ambiente de multitrayectoria "dinámico" en el que el receptor está en movimiento y las dos versiones de la señal de banda angosta transmitida llegan al

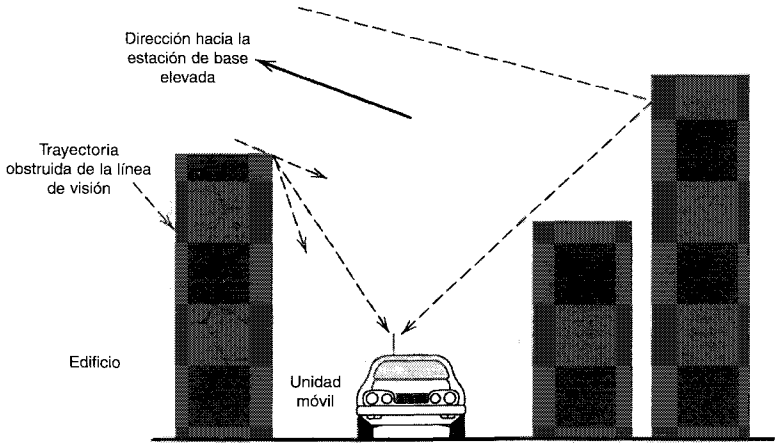


FIGURA 8.13 Ilustración del mecanismo de la propagación de radio en áreas urbanas. (Tomado de Parsons con autorización.)

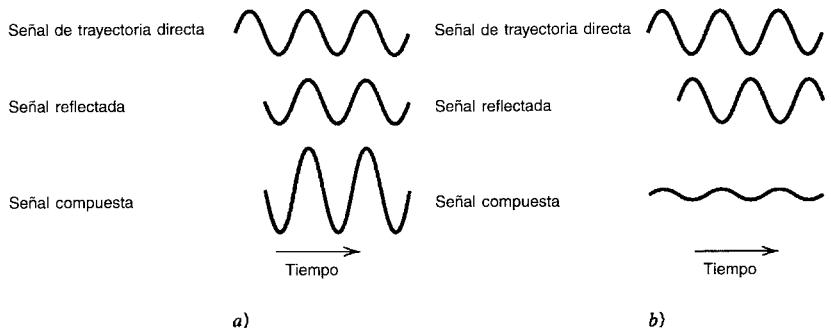


FIGURA 8.14 *a*) Formas constructivas, y *b*) destructivas del fenómeno de multitrayectoria para señales senoidales.

receptor por trayectorias de longitudes diferentes. Debido al movimiento del receptor, hay un cambio continuo en la longitud de cada trayectoria de propagación. Por tanto, el corrimiento de fase relativo entre las dos componentes de la señal recibida es una función de la ubicación espacial del receptor. Cuando este último se mueve, encontramos que la amplitud (envolvente) recibida ya no es constante como lo fue en el caso de un ambiente estático; más bien, varía con la distancia en la forma que se ilustra en la figura 8.16. En la parte superior de esta figura hemos incluido también las relaciones de fasores para las dos componentes de la señal recibida en diversas localizaciones del receptor. La figura 8.16 muestra que hay una adición constructiva en algunas ubicaciones, y la casi cancelación completa en algunas otras. Este fenómeno se conoce como el *desvanecimiento de la señal.*

En un ambiente de radio móvil encontrado en la práctica, pueden existir desde luego una multitud de trayectorias de propagación con diferentes longitudes, y sus contribuciones a la señal

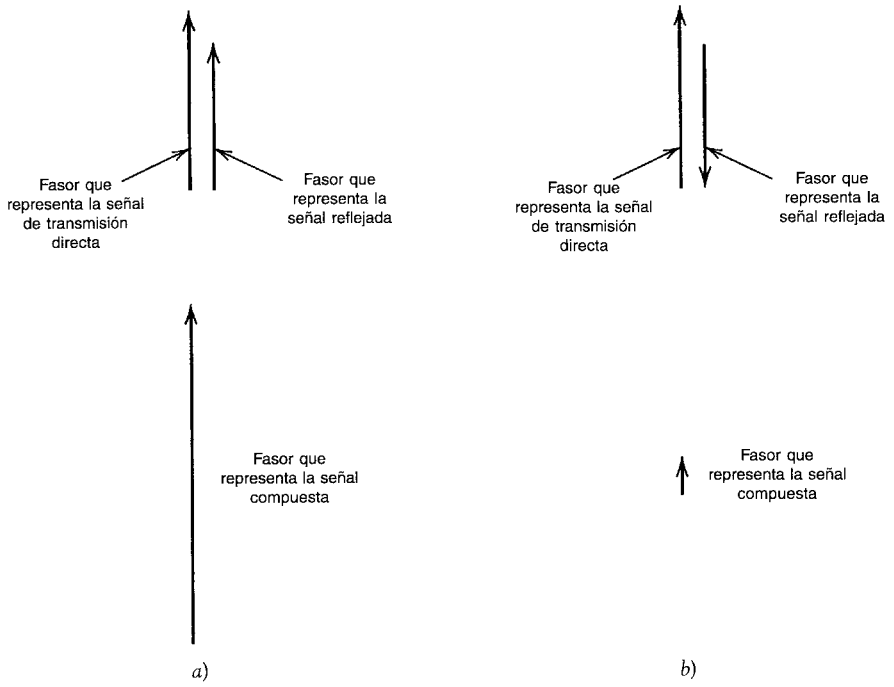


FIGURA 8.15 Representación de fasores de *a*) formas constructivas y *b*) destructivas de multitrayectoria.

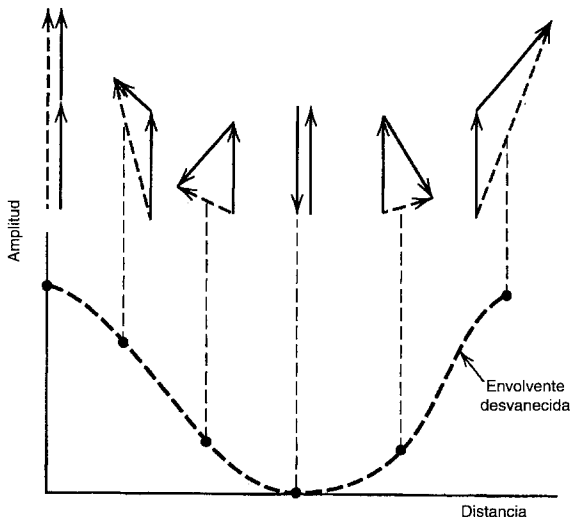


FIGURA 8.16 Ilustración de cómo la envolvente se desvanece cuando dos señales entrantes se combinan con diferentes fases. (Tomado de Parsons con autorización.)

recibida podrían combinarse en una diversidad de maneras. El resultado neto es que la envolvente de la señal recibida varía con la ubicación de una manera complicada, como se indica por medio del registro experimental de la envolvente de la señal recibida en un área urbana que se presenta en la figura 8.17. Esta figura exhibe claramente la naturaleza de desvanecimiento de la señal recibida. La envolvente de la señal recibida en la figura 8.17 se mide en dBm. La unidad dBm se define como $10 \log_{10}(P/P_0)$, con P denotando la potencia que se está midiendo y $P_0 = 1$ miliwatt. En el caso de la figura 8.17, P es la potencia instantánea en la envolvente de la señal recibida.

El desvanecimiento de la señal es en esencia un *fenómeno espacial* que se manifiesta a sí mismo en el dominio del tiempo cuando el receptor se mueve. Estas variaciones pueden relacionarse con el movimiento del receptor del modo siguiente. De manera específica, considere la situación que se ilustra en la figura 8.18, donde se supone que el receptor se está moviendo a lo largo de la línea AA' con una velocidad constante v. Se considera también que la señal recibida

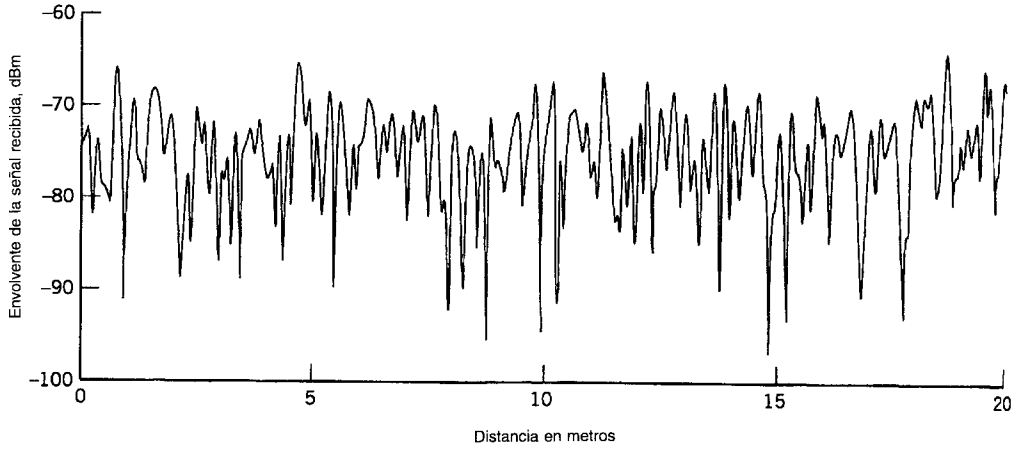


FIGURA 8.17 Registro experimental de la envolvente de la señal recibida en un área urbana. (Tomada de Parsons con autorización.)

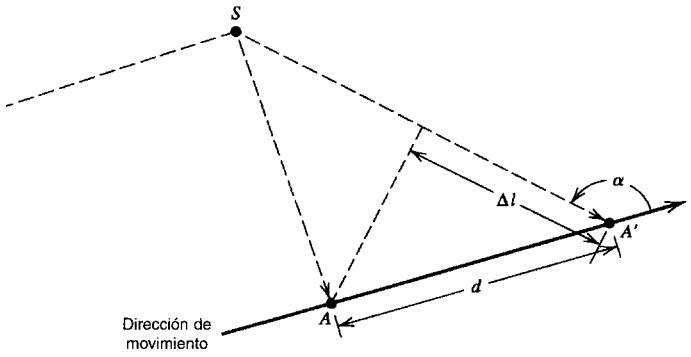


FIGURA 8.18 Ilustración del cálculo del corrimiento Doppler.

se debe a una onda de radio que proviene de un dispersor denominado S. Dejemos que Δt denote el tiempo que tarda el receptor en moverse del punto A a A'. Utilizando la notación que se describe en la figura 8.18, se deduce que el cambio incremental en la longitud de la trayectoria de la onda de radio corresponde a

$$\begin{aligned} \Delta l &= d \cos \alpha \\ &= -v \, \Delta t \cos \alpha \end{aligned} \tag{8.34}$$

donde α es el ángulo espacial entre la onda de radio entrante y la dirección del movimiento del receptor. De forma correspondiente, el cambio en el ángulo de fase de la señal recibida en el punto A' con respecto al punto A está dado por

$$\begin{aligned} \Delta\phi &= \frac{2\pi}{\lambda} \Delta l \\ &= -\frac{2\pi v \, \Delta t}{\lambda} \cos \alpha \end{aligned} \tag{8.35}$$

donde λ es la longitud de onda de la señal de radio. El cambio manifiesto en la frecuencia, o el *corrimiento Doppler*, es entonces

$$\begin{aligned} v &= \frac{1}{2\pi} \frac{\Delta\phi}{\Delta t} \\ &= -\frac{v}{\lambda} \cos \alpha \end{aligned} \tag{8.36}$$

El corrimiento Doppler v es positivo (dando lugar a un aumento de frecuencia) cuando las ondas de radio arriban provenientes del frente de la unidad móvil, y es negativo cuando las ondas de radio arriban provenientes de la parte trasera de la unidad móvil.

8.6 *Caracterización estadística de canales de multitrayectoria*

La caracterización de banda angosta del ambiente de multitrayectoria que se describe en la sección 8.5 resulta apropiada en las transmisiones de radio móvil donde el ancho de banda de la señal

es muy pequeño en comparación con el recíproco de la dispersión en los retardos de la trayectoria de propagación. La multitrayectoria en un ambiente de este tipo produce dos efectos: el rápido desvanecimiento de la envolvente de la señal recibida y una dispersión de los corrimientos Doppler en el espectro recibido. Las señales de la vida real radiadas en un ambiente de radio móvil pueden ocupar un ancho de banda suficientemente amplio para requerir consideraciones más detalladas de los efectos de la propagación multitrayectoria sobre la señal recibida. En esta sección presentamos una caracterización estadística de un canal de radio móvil.*

Considere un canal de radio móvil con múltiples trayectorias de propagación. De acuerdo con la notación compleja descrita en el apéndice 2, es posible expresar la señal de banda de paso como

$$s(t)\,\mathrm{Re}[\tilde{s}(t)\exp(j2\pi f_c t)] \tag{8.37}$$

donde $\tilde{s}(t)$ es la envolvente compleja (pasobajas) de $s(t)$ y f_c es una frecuencia de portadora nominal. Puesto que el canal varía en el tiempo debido a los efectos de multitrayectoria, la respuesta al impulso del mismo depende del retardo y, por tanto, es una función variable en el tiempo. Dejemos que la respuesta al impulso del canal se exprese como

$$h(t)\,\mathrm{Re}[\tilde{s}(t)\exp(j2\pi f_c t)] \tag{8.38}$$

donde $\tilde{h}(\tau;t)$ es la respuesta al impulso compleja (pasobajas) del canal y τ es la variable de retardo. La respuesta al impulso compleja $\tilde{h}(\tau;t)$ recibe el nombre de *función de dispersión del retardo de la entrada* del canal. La envolvente compleja (pasobajas) de la salida del canal está definida por la integral de convolución

$$\tilde{s}_0(t) = \frac{1}{2}\int_{-\infty}^{\infty} \tilde{s}(t-\tau)\tilde{h}(\tau;t)\,d\tau \tag{8.39}$$

donde el factor de escalamiento $\frac{1}{2}$ es el resultado de utilizar notación compleja.

En general, el comportamiento de un canal de radio móvil puede describirse sólo en términos estadísticos. Para los propósitos analíticos, esta función de dispersión del retardo $\tilde{h}(\tau;t)$ puede modelarse como un proceso gaussiano de valores complejos y media cero. En ese caso, en cualquier tiempo t la envolvente $|\tilde{h}(\tau;t)|$ presenta una distribución de Rayleigh y el canal se conoce como un *canal de desvanecimiento de Rayleigh*. Sin embargo, cuando el ambiente de radio móvil incluye dispersores *fijos*, ya no es justificable utilizar el modelo de media cero para describir la función de dispersión del retardo de la entrada $\tilde{h}(\tau;t)$. En tal caso, resulta más apropiado utilizar la distribución de Rician para describir la envolvente $|\tilde{h}(\tau;t)|$, y el canal se conoce como *canal de desvanecimiento de Rician*. Las distribuciones de Rayleigh y Rician para un proceso aleatorio de valores reales se consideraron en el capítulo 1. En la exposición que se presenta en este capítulo sólo nos ocuparemos del canal de desvanecimiento de Rayleigh.

La *función de transferencia variante con el tiempo* del canal se define como la transformada de Fourier de la función de dispersión del retardo de la entrada $\tilde{h}(\tau;t)$ con respecto a la variable de retardo τ, como indica

$$\tilde{H}(f;t) = \int_{-\infty}^{\infty} \tilde{h}(\tau;t)\exp(-j2\pi f\tau)\,d\tau \tag{8.40}$$

donde f denota la variable de frecuencia. La función de transferencia variante con el tiempo $\tilde{H}(f;t)$ puede considerarse como una característica de transmisión de frecuencia del canal.

*Los lectores que no estén interesados en los detalles matemáticos relativos a la caracterización estadística de los canales de multitrayectoria con desvanecimiento, tienen la opción de omitir el material que se presenta en esta sección, excepto la subsección sobre la clasificación de los canales multitrayectoria.

Para una caracterización estadística del canal, hacemos las siguientes suposiciones:

- La función de dispersión de retardo de la entrada $\tilde{h}(\tau;t)$ es un *proceso gaussiano de valores complejos y media cero*. Nuestro interés se restringe al desvanecimiento de corto plazo; por tanto, resulta razonable suponer que $\tilde{h}(\tau;t)$ también es *estacionaria*. A causa de que la transformación de Fourier es lineal, la función de transferencia variante con el tiempo $\tilde{H}(f;t)$ tiene estadística similar.
- El canal es de *dispersión no correlacionada*, lo que quiere decir que las aportaciones de los dispersores con diferentes retardos de propagación son no correlacionadas.

Considere entonces la función de autocorrelación de la función de dispersión del retardo de entrada $\tilde{h}(\tau;t)$. Puesto que $\tilde{h}(\tau;t)$ es de valor complejo, empleamos la siguiente definición para la función de autocorrelación:

$$R_{\tilde{h}}(\tau_1,t_1;\tau_2,t_2)=E\left[\tilde{h}^*(\tau_1;t_1)\tilde{h}(\tau_2;t_2)\right] \tag{8.41}$$

donde E es el operador de esperanza estadística, el asterisco denota conjugación compleja, τ_1 y τ_2 son los retardos de la propagación de las dos trayectorias implicadas en el cálculo y t_1 y t_2 son los tiempos en los cuales se observan las salidas de las dos trayectorias. Apelando a la estacionariedad de la variable de tiempo t y a la dispersión no correlacionada en la variable de retardo de tiempo τ, es posible reformular la función de autocorrelación de $\tilde{h}(\tau;t)$ como

$$\begin{aligned} R_{\tilde{h}}(\tau_1,\tau_2;\Delta t) &= E\left[\tilde{h}^*(\tau_1;t)\tilde{h}(\tau_2;t+\Delta t)\right] \\ &= r_{\tilde{h}}(\tau_1;\Delta t)\ \delta(\tau_1-\tau_2) \end{aligned} \tag{8.42}$$

donde Δt es la diferencia entre los tiempos de observación, y $\delta(\tau_1 - \tau_2)$ es una función delta. Utilizando τ en lugar de τ_1, la función restante en la ecuación (8.42) se redefine como

$$r_{\tilde{h}}(\tau;\Delta t)=E\left[\tilde{h}(\tau;t)\tilde{h}^*(\tau;t+\Delta t)\right] \tag{8.43}$$

La función $r_{\tilde{h}}(\tau;\Delta t)$ recibe el nombre de *perfil de autocorrelación* de multitrayectoria del canal.

Considere a continuación una caracterización estadística del canal en términos de la función de transferencia variante con el tiempo de valor complejo $\tilde{H}(f;t)$. Siguiendo una formulación similar a la que se describió en la ecuación (8.41), la función de autocorrelación $\tilde{H}(f;t)$ está definida por

$$R_{\tilde{H}}(f_1,t_1;f_2,t_2)=E\left[\tilde{H}^*(f_1;t_1)\tilde{H}(f_2;t_2)\right] \tag{8.44}$$

donde f_1 y f_2 representan dos frecuencias en el espectro de la señal transmitida. La función de autocorrelación $R_{\tilde{H}}(f_1,t_1;f_2,t_2)$ proporciona una medida estadística del grado al cual la señal es distorsionada por la transmisión a través del canal. A partir de las ecuaciones (8.40), (8.41) y (8.44) encontramos que las funciones de autocorrelación $R_{\tilde{H}}(f_1,t_1;f_2,t_2)$ y $R_{\tilde{h}}(\tau_1,t_1;\tau_2,t_2)$ se relacionan por medio de una forma de la transformación de Fourier bidimensional de la manera que sigue:

$$R_{\tilde{H}}(f_1,t_1;f_2,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}R_{\tilde{h}}(\tau_1,t_1;\tau_2,t_2)\ \exp\left[j2\pi(f_1\tau_1-f_2\tau_2)\right]\,d\tau_1\ d\tau_2 \tag{8.45}$$

Apelando a la estacionariedad en el dominio del tiempo, es posible reformular la ecuación (8.44) como

$$R_{\tilde{H}}(f_1,f_2;\Delta t)=E\left[\tilde{H}^*(f_1;t)\tilde{H}(f_2;t+\Delta t)\right] \tag{8.46}$$

Esta definición sugiere que la función de autocorrelación $R_{\tilde{H}}(f_1,f_2;\Delta t)$ puede medirse por pares de tonos espaciados para efectuar mediciones de correlación cruzada sobre las salidas resul-

tantes del canal. Una medición de estas características presupone estacionariedad en el dominio del tiempo. Si también suponemos estacionariedad en el dominio de la frecuencia, podemos ir un paso adelante y escribir

$$R_{\widetilde{H}}(f, f+\Delta f;\Delta t) = r_{\widetilde{H}}(\Delta f;\Delta t) = E\left[\widetilde{H}^*(f;t)\widetilde{H}(f+\Delta f;t+\Delta t)\right] \tag{8.47}$$

Esta forma especializada de la función de autocorrelación de $\widetilde{H}(f;t)$ es de hecho la transformada de Fourier del perfil de autocorrelación de multitrayectoria $r_{\widetilde{h}}(\tau;\Delta t)$ con respecto a la variable del tiempo de retardo τ, según expresa

$$r_{\widetilde{H}}(\Delta f;\Delta t) = \int_{-\infty}^{\infty} r_{\widetilde{h}}(\tau;\Delta t)\ \exp(-j2\pi\tau\ \Delta f)\ d\tau \tag{8.48}$$

La función $r_{\widetilde{H}}(\Delta f;\Delta t)$ se denomina la *función de correlación de tiempo espaciado frecuencia espaciada* del canal.

Por último introducimos una función $S(\tau;v)$ que forma un par de transformadas de Fourier con el perfil de autocorrelación multitrayectoria $r_{\widetilde{h}}(\tau;\Delta t)$ con respecto a la variable Δt, según indica

$$S(\tau;v) = \int_{-\infty}^{\infty} r_{\widetilde{h}}(\tau;\Delta t)\ \exp(-j2\pi v\ \Delta t)\ d(\Delta t) \tag{8.49}$$

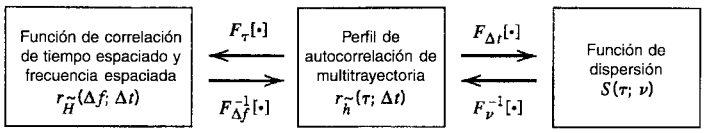
y

$$r_{\widetilde{h}} = (t;\Delta t) = \int_{-\infty}^{\infty} S(\tau;v)\ \exp(j2\pi v\ \Delta t)\ dv \tag{8.50}$$

La función $S(\tau;v)$ puede definirse también en términos de $r_{\widetilde{H}}(\Delta f;\Delta t)$ aplicando una forma de una doble transformación de Fourier: una transformada de Fourier con respecto a la variable de tiempo Δt y una transformada inversa de Fourier con respecto a la variable de frecuencia Δf. Es decir,

$$S(\tau;v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_{\widetilde{H}}(\Delta f;\Delta t)\ \exp(-j2\pi v\,\Delta t)\ \exp(j2\pi\tau\,\Delta f)\ d(\Delta t)\ d(\Delta f) \tag{8.51}$$

La figura 8.19 presenta las relaciones funcionales entre $r_{\widetilde{h}}(\tau;\Delta t)$, $r_{\widetilde{H}}(\Delta f;\Delta t)$ y $S(\tau;v)$ en términos de la transformada de Fourier y su inversa.



$F_\tau[\cdot]$: Transformada de Fourier con respecto al retardo τ

$F_{\Delta f}^{-1}[\cdot]$: Transformada inversa de Fourier con respecto al incremento de frecuencia Δf

$F_{\Delta t}[\cdot]$: Transformada de Fourier con respecto al incremento del tiempo Δt

$F_v^{-1}[\cdot]$: Transformada inversa de Fourier con respecto al corrimiento Doppler v

FIGURA 8.19 Relaciones funcionales entre el perfil de autocorrelación de multitrayectoria $r_{\widetilde{h}}(\tau;\Delta t)$, la función de correlación de tiempo espaciado y frecuencia espaciada $r_{\widetilde{H}}(\Delta f;\Delta t)$, y la función de dispersión $S(\tau;v)$.

La función $S(\tau;v)$ se conoce como la *función dispersora* del canal. En lo referente a una interpretación física de ésta, considere la transmisión de un tono sencillo de frecuencia f' (relativa a la portadora). La envolvente compleja de la salida del filtro resultante es

$$\tilde{s}_o(t) = \exp(j2\pi f' t)\tilde{H}(f';t) \tag{8.52}$$

La función de autocorrelación de $\tilde{s}_o(t)$ es

$$\begin{aligned} E\left[\tilde{s}_o^*(t)\tilde{s}_o(t+\Delta t)\right] &= \exp(j2\pi f'\Delta t)E\left[\tilde{H}^*(f';t)\tilde{H}(f';t+\Delta t)\right] \\ &= \exp(j2\pi f'\Delta t)r_{\tilde{H}}(0;\Delta t) \end{aligned} \tag{8.53}$$

donde, en la última línea, hemos utilizado la ecuación (8.47). Al dejar $\Delta f = 0$ en la ecuación (8.48) y utilizar después la ecuación (8.50), es posible escribir

$$\begin{aligned} r_{\tilde{H}}(0;Dt) &= \int_{-\infty}^{\infty} r_{\tilde{h}}(\tau;\Delta t)\, d\tau \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} r_{\tilde{h}}S(\tau;v)\, d\tau\right]\exp(j2\pi v\,\Delta t)\, dv \end{aligned} \tag{8.54}$$

Por tanto, podemos considerar la integral

$$\int_{-\infty}^{\infty} S(t;v)\, d\tau$$

como la densidad espectral de potencia de la salida del canal relativa a la frecuencia f' del tono transmitido, y con el corrimiento Doppler v actuando como la variable de frecuencia. Generalizando este resultado, podemos establecer que la función dispersora $S(\tau;v)$ proporciona una medida estadística de la potencia de salida del canal, expresada como una función de retardo de tiempo τ y del corrimiento Doppler v.

■ Dispersión del retardo y dispersión Doppler

Dejando $\Delta t = 0$ en la ecuación (8.43), es posible escribir

$$\begin{aligned} P_{\tilde{h}}(\tau) &= r_{\tilde{h}}(\tau;0) \\ &= E\left[\left|\tilde{h}(\tau;t)\right|^2\right] \end{aligned} \tag{8.55}$$

La función $P_{\tilde{h}}(\tau)$ describe la intensidad (promediada sobre las fluctuaciones de desvanecimiento) del proceso dispersor en el retardo de propagación τ. Por tanto, $P_{\tilde{h}}(\tau)$ recibe el nombre de *espectro de potencia de retardo* o *perfil de intensidad de multitrayectoria* del canal. El espectro de la potencia de retardo puede definirse también en términos de la función dispersora $S(\tau;\, v)$ al promediar sobre todos los corrimientos Doppler. De modo específico, dejando $\Delta t = 0$ en la ecuación (8.50) y utilizando luego la primera línea de la ecuación (8.55), podemos escribir

$$P_{\tilde{h}}(\tau) = \int_{-\infty}^{\infty} S(t;v)\, dv \tag{8.56}$$

La figura 8.20 muestra un ejemplo de un espectro de la potencia de retardo que describe una gráfica representativa de la densidad espectral de potencia en función de retardo de exceso; este último se mide con respecto al retardo del tiempo para la trayectoria de eco más corta. Advierta, como en la figura 8.17, que la potencia se mide en dBm. El "nivel de umbral" incluido en la figura 8.20 define el nivel de potencia debajo del cual el receptor deja de operar de modo satisfactorio.

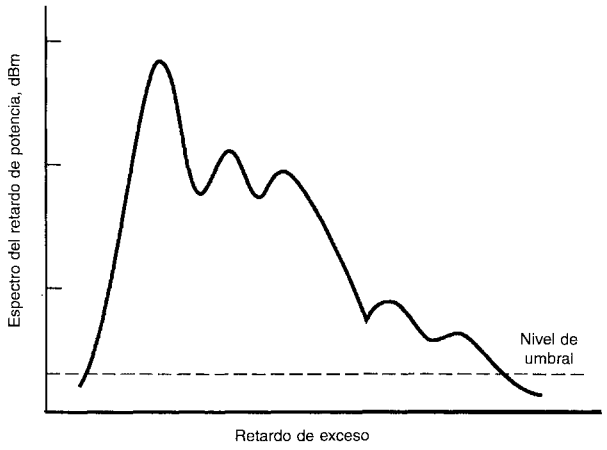


FIGURA 8.20 Ejemplo de un perfil de retardo de potencia para un canal de radio móvil. (Tomado de Parsons con autorización.)

Dos momentos estadísticos de $P_{\tilde{h}}(\tau)$ de interés son el *retardo promedio*, τ_{pro}, y la *dispersión de retardo*, σ_τ. El retardo promedio se define como el primer momento central (es decir, la media) de $P_{\tilde{h}}(\tau)$, de acuerdo a como señala

$$\tau_{pro} = \frac{\int_0^\infty \tau P_{\tilde{h}}(\tau)\, d\tau}{\int_0^\infty P_{\tilde{h}}(\tau)\, d\tau} \tag{8.57}$$

La dispersión del retardo se define como la raíz cuadrada del segundo momento central de $P_{\tilde{h}}(\tau)$:

$$\sigma_\tau = \left(\frac{\int_0^\infty (\tau - \tau_{pro})^2 P_{\tilde{h}}(\tau)\, d\tau}{\int_0^\infty P_{\tilde{h}}(\tau)\, d\tau} \right)^{1/2} \tag{8.58}$$

El recíproco de la dispersión del retardo σ_τ es una medida del *ancho de banda de coherencia* del canal, el cual se denota mediante B_c.

Considere en seguida la cuestión de relacionar los efectos Doppler con las variaciones de tiempo del canal. Para este propósito, fijamos primero $\Delta f = 0$, lo cual corresponde a la transmisión de un solo tono (de alguna frecuencia apropiada) por el canal. La función de correlación de tiempo espaciado frecuencia espaciada del canal se reduce entonces a $r_{\tilde{H}}(0;\Delta t)$. Por tanto, al evaluar la transformada de Fourier de esta función con respecto a la variable de tiempo Δt, podemos escribir

$$S_{\tilde{H}}(v) = \int_{-\infty}^{\infty} r_{\tilde{H}}(0;\Delta t)\, \exp(-j2\pi v\, \Delta t)\, d(\Delta t) \tag{8.59}$$

La función $S_{\tilde{H}}(v)$ define el espectro de potencia de la salida del canal expresado como una función del corrimiento Doppler v; por tanto, recibe el nombre de *espectro Doppler* del canal. El

espectro Doppler puede definirse también en términos de la función dispersora al promediarlo sobre todos los retardos de la propagación, de acuerdo con

$$S_{\tilde{H}}(\nu) = \int_{-\infty}^{\infty} S(\tau;\nu)\, d\tau \tag{8.60}$$

El corrimiento Doppler ν puede tomar valores positivos y negativos con igual verosimilitud. Por tanto, el corrimiento Doppler medio es igual a cero. La raíz cuadrada del segundo momento del espectro Doppler está definida entonces por

$$\sigma_{\nu} = \left(\frac{\int_{0}^{\infty} \nu^2 S_{\tilde{H}}(\nu)\, d\nu}{\int_{0}^{\infty} S_{\tilde{H}}(\nu)\, d\nu} \right)^{1/2} \tag{8.61}$$

El parámetro $\sigma\nu$ proporciona una medida del ancho del espectro Doppler; en consecuencia, recibe el nombre de *dispersión Doppler* del canal. El recíproco de la dispersión Doppler se llama *tiempo de coherencia* del canal, el cual se denota por medio de τ_c.

Otro parámetro útil que se usa a menudo en las mediciones es la *tasa de desvanecimiento* del canal. En un canal de desvanecimiento Rayleigh la tasa de desvanecimiento promedio se relaciona con la dispersión Doppler $\sigma\nu$ como

$$f_e = 1.475\sigma\nu \quad \text{cruces por segundo} \tag{8.62}$$

Como su nombre lo indica, la tasa de desvanecimiento proporciona una medida de la rapidez del desvanecimiento del canal.

Algunos valores típicos que se encuentran en el ambiente del radio móvil son los siguientes:

- La dispersión de retardo, σ_τ, asciende hasta 20 μs.
- La dispersión Doppler, σ_ν, debida al movimiento de un vehículo, puede extenderse hasta 40–80 Hz.

Clasificación de canales multitrayectoria

La forma particular de desvanecimiento experimentada por un canal multitrayectoria depende de si la caracterización del canal se considera en el dominio de la frecuencia o en el dominio del tiempo.

Cuando el canal se considera en el dominio de la frecuencia, el parámetro de interés es el ancho de banda de coherencia del canal, B_c, que es una medida del ancho de banda de transmisión para la cual la distorsión de la señal a través del canal se vuelve observable. Se dice que un canal multitrayectoria será *selectivo de frecuencia* si el ancho de banda de coherencia del canal es pequeño comparado con el ancho de banda de la señal transmitida. En tal situación, el canal tiene un efecto de filtrado en el sentido de que las componentes senoidales, con una separación de frecuencia superior que el ancho de banda de coherencia del canal, se tratan de manera diferente. No obstante, si el ancho de banda de coherencia del canal es grande en comparación con el ancho de banda del mensaje, se dice que el desvanecimiento será *no selectivo de frecuencia* o *plano de frecuencia*.

Cuando el canal se considera en el dominio del tiempo, el parámetro de interés es el tiempo de coherencia, τ_c, el cual ofrece una medida de la duración de la señal transmitida para la cual la distorsión a través del canal se vuelve observable. Se dice que el desvanecimiento será *selectivo de tiempo* si el tiempo de coherencia del canal es pequeño en comparación con la duración de la señal recibida (es decir, el tiempo para el cual la señal está transmitiéndose). En la transmisión digital, la duración de la señal recibida se considera como la duración del símbolo más la dispersión del retardo del canal. Si, no obstante, el tiempo de coherencia del canal es grande en comparación con la duración de la señal recibida, se dice que el desvanecimiento es

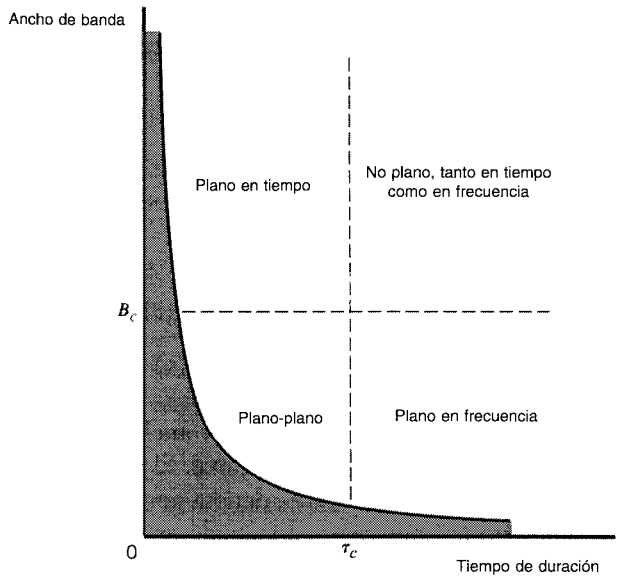


FIGURA 8.21 Ilustración de las cuatro clases de canales multitrayectoria: τ_c = tiempo coherente, B_c = ancho de banda coherente.

no selectivo de tiempo o *plano de tiempo*, en el sentido de que el canal aparece invariante con el tiempo para la señal transmitida.

A la luz de esta explicación, es posible clasificar los canales multitrayectoria de la manera siguiente:

- *Canal plano–plano*, que es plano tanto en frecuencia como en tiempo.
- *Canal plano de frecuencia*, que es plano exclusivamente en frecuencia.
- *Canal plano de tiempo*, que es plano sólo en el tiempo.
- *Canal no plano*, que no es plano ni en la frecuencia ni en el tiempo. Un canal de este tipo se denomina algunas veces *canal doblemente disperso*.

La clasificación de los canales multitrayectoria, con base en este método, se muestra en la figura 8.21. El área prohibida, que se muestra sombreada en esta figura, sigue de la relación inversa que existe entre el ancho de banda y la duración del tiempo.

8.7 *Señalización binaria sobre un canal de desvanecimiento de Rayleigh*

En el capítulo 6 determinamos la probabilidad promedio del error de símbolo en la transmisión de datos binarios por un canal corrompido por ruido blanco gaussiano aditivo. En un ambiente de radio móvil tenemos que considerar un efecto adicional, a saber, las fluctuaciones en la amplitud y la fase de la señal recibida que resultan de los efectos de multitrayectoria. Específicamente, considere la transmisión de datos binarios por un canal de desvanecimiento de Rayleigh, para el cual la envolvente compleja (pasobanda) de la señal recibida se modela de la forma siguiente:

$$\tilde{x}(t) = \alpha \exp(-j\phi)\tilde{s}(t) + \tilde{w}(t) \tag{8.63}$$

donde $\tilde{s}(t)$ es la envolvente compleja de la señal transmitida (pasobanda), α es una variable aleatoria con distribución de Rayleigh que describe la atenuación en la transmisión, ϕ es una

variable aleatoria distribuida uniformemente que describe el corrimiento de fase en la transmisión, y $\tilde{w}(t)$ es el proceso de ruido blanco gaussiano de valores complejos. Se supone que el canal es plano tanto en tiempo como en frecuencia, de modo que podemos estimar el corrimiento de fase ϕ a partir de la señal recibida sin error. Supóngase entonces que el corrimiento de fase por llaveo binaria coherente se usa para realizar la transmisión de datos. Bajo la condición de que α es fija o constante sobre un intervalo de bit, es posible adaptar la ecuación (6.20) del capítulo 6 a la situación presente expresando la probabilidad promedio del error de símbolo (es decir, la tasa de error de bits) debida al ruido blanco gaussiano aditivo actuando exclusivamente:

$$P_e(\gamma) = \tfrac{1}{2}\operatorname{ferc}(\sqrt{\gamma}) \tag{8.64}$$

donde γ es una versión atenuada de la relación de la energía de la señal transmitida por bit a la densidad espectral de ruido E_b/N_0, de acuerdo a como indica

$$\gamma = \frac{\alpha^2 E_b}{N_0} \tag{8.65}$$

Ahora bien, con respecto a lo que se refiere al canal de radio móvil, es posible considerar $P_e(\gamma)$ como una probabilidad condicional dado que α es fija. De este modo, para evaluar la probabilidad promedio del error de símbolo en la presencia combinada del desvanecimiento y el ruido, debemos promediar $P_e(\gamma)$ sobre todos los valores posibles de gama, en la forma indicada por

$$P_e = \int_0^\infty P_e(\gamma) f(\gamma)\, d\gamma \tag{8.66}$$

donde $f(\gamma)$ es la función de densidad de probabilidad de γ. De acuerdo con la ecuación (8.65) advertimos que γ depende del valor al cuadrado de α. Puesto que α presenta una distribución de Rayleigh, determinamos que γ tiene una *distribución ji cuadrada* con dos grados de libertad.[7] En particular, podemos expresar la función de densidad de probabilidad de γ como

$$f(\gamma) = \frac{1}{\gamma_0}\exp\left(-\frac{\gamma}{\gamma_0}\right), \quad \gamma \geq 0 \tag{8.67}$$

El término γ_0 es el *valor medio de la relación de la energía de la señal recibida por bit a la densidad espectral del ruido*, la cual se define mediante

$$\begin{aligned}\gamma_0 &= E[\gamma] \\ &= \frac{E_b}{N_0} E[\alpha^2]\end{aligned} \tag{8.68}$$

donde $E[\alpha^2]$ es el valor cuadrático medio de la variable aleatoria α con distribución de Rayleigh. Al sustituir las ecuaciones (8.64) y (8.67) en la (8.66), y efectuar la integración, obtenemos el resultado final

$$P_e = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma_0}{1+\gamma_0}}\right) \tag{8.69}$$

La ecuación (8.69) define la tasa de error de bits para el corrimiento de fase por llaveo (PSK) binaria coherente por un canal de desvanecimiento de Rayleigh plano-plano. Siguiendo un método similar, es posible obtener las tasas de error de bits correspondientes para el corrimiento de frecuencia por llaveo (FSK) binaria coherente, el corrimiento de fase por llaveo diferencial (DPSK) y la FSK binaria no coherente. Los resultados de estas evaluaciones se resumen en la tabla 8.2. En la figura 8.22 hemos utilizado las fórmulas exactas de la tabla 8.2 para graficar la tasa de error de bits en función de γ_0 expresada en decibeles. Por el bien de la comparación, hemos incluido también en la figura 8.22 gráficas de tasas de error de bits de la PSK binaria coherente y la FSK binaria no coherente en un canal sin desvanecimiento. Observamos que el desvanecimiento de Rayleigh produce una seria degradación en el desempeño ante el

TABLA 8.2 *Tasas de error de bits para transmisión de señales binarias sobre un canal de desvanecimiento de Rayleigh plano-plano*

Tipo de transmisión de señal	*Fórmula exacta para la tasa de error de bits* P_e	*Fórmula aproximada para la tasa de error de bits, suponiendo un valor de* γ_0 *grande*
PSK binaria coherente	$\frac{1}{2}\left(1-\sqrt{\frac{\gamma_0}{1+\gamma_0}}\right)$	$\frac{1}{4\gamma_0}$
FSK binaria coherente	$\frac{1}{2}\left(1-\sqrt{\frac{\gamma_0}{2+\gamma_0}}\right)$	$\frac{1}{2\gamma_0}$
DPSK binaria	$\frac{1}{2(1+\gamma_0)}$	$\frac{1}{2\gamma_0}$
FSK binaria no coherente	$\frac{1}{2+\gamma_0}$	$\frac{1}{\gamma_0}$

ruido de un sistema de transmisión pasobanda digital, midiéndose la degradación en decenas de decibeles de la relación señal a ruido media adicional comparada con un canal sin desvanecimiento para la misma tasa de error de bits. En particular, para γ_0 grande es posible obtener las fórmulas aproximadas dadas en la última columna de la tabla 8.2, de acuerdo con las cuales la disminución asintótica en la tasa de error de bits con la relación de energía promedio de la señal por bit a la densidad espectral de ruido γ_0 sigue una *ley inversa*. Este comportamiento es sumamente diferente del caso de un canal sin desvanecimiento, en el cual la reducción asintótica en la tasa de error de bits con γ_0 sigue una *ley exponencial*.

La implicación práctica de esta diferencia es que en un ambiente de radio móvil tenemos que proporcionar un incremento considerable en la relación señal a ruido media (relativa a un ambiente sin desvanecimiento), a fin de asegurar una tasa de error de bits que es suficientemente baja para uso práctico. Para conseguir tal requerimiento, necesitamos aumentar la potencia transmitida, el tamaño de la antena, etc., lo cual puede resultar costoso en términos de la puesta en práctica. De manera alternativa, es posible utilizar técnicas de modulación y recepción especiales que son menos vulnerables a los efectos del desvanecimiento. Entre dichas técnicas, las más conocidas y las que se usan con mayor amplitud son las técnicas combinadas de receptor múltiple conocidas colectivamente como *diversidad*, de la cual se presenta una breve explicación en lo que sigue.

■ TÉCNICAS DE DIVERSIDAD

La diversidad puede considerarse como una forma de redundancia. En particular, si es posible transmitir varias réplicas de la señal de mensaje de manera simultánea por canales con desvanecimiento independientes, entonces hay una buena verosimilitud de que al menos una de las señales recibidas no será degradada en forma considerable por el desvanecimiento. Existen varios métodos para llevar a cabo una provisión de este tipo en el contexto de nuestra explicación presente; las siguientes técnicas de diversidad son de interés particular:

- Diversidad de frecuencia
- Diversidad de tiempo (repetición de la señal)
- Diversidad del espacio

En la *diversidad de frecuencia* la señal de mensaje se transmite utilizando varias portadoras que están lo suficientemente espaciadas una de otra para ofrecer versiones con desvanecimien-

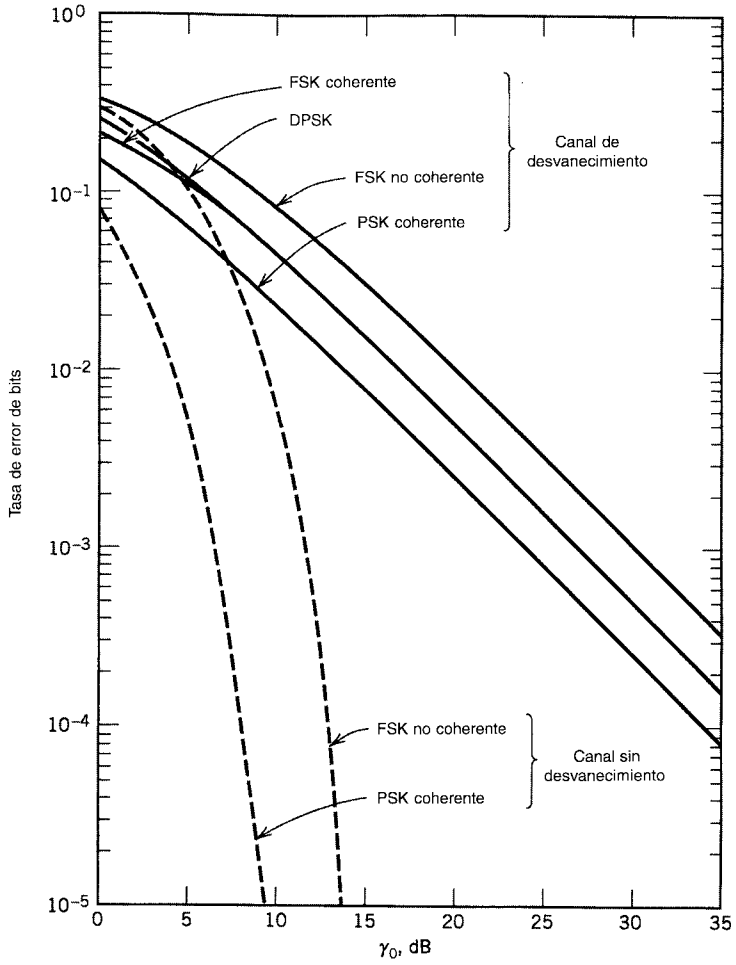


FIGURA 8.22 Desempeño de esquemas de transmisión de señales binarias por un canal de desvanecimiento de Rayleigh, mostrado como curvas continuas; las curvas discontinuas pertenecen al canal sin desvanecimiento.

to independientes de la señal. Esto puede lograrse eligiendo un espaciamiento de frecuencia igual o mayor que el ancho de banda de coherencia del canal.

En la *diversidad de tiempo* la misma señal de mensaje se transmite en diferentes franjas de tiempo, con el espaciamiento entre franjas sucesivas igual o mayor que el tiempo de coherencia del canal. La diversidad de tiempo también puede asemejarse al uso del código de repetición para la codificación de control de errores. (La codificación del control de errores se explica en el capítulo 10.)

En la *diversidad de espacio* se usan múltiples antenas transmisoras o receptoras (o de ambos tipos), con el espaciamiento entre antenas adyacentes eligiéndose de manera que se asegure la independencia de los eventos de desvanecimiento; esto puede satisfacerse espaciando las antenas adyacentes por al menos siete veces la longitud de onda de radio.

Debido a que mediante uno de estos medios creamos L canales de desvanecimiento independiente, es posible utilizar entonces una *estructura de combinación de diversidad lineal* que incluya L receptores separados, como se describe en la figura 8.23. El sistema se diseña para compensar únicamente los *efectos de corto plazo* de un canal de desvanecimiento. Además, se supone que las *estimaciones sin ruido* de los factores de atenuación del canal $\{\alpha_\ell\}$ y los corrimientos de fase del canal $\{\phi_\ell\}$ se tienen a mano. En ese caso, el combinador lineal alcanza un desempe-

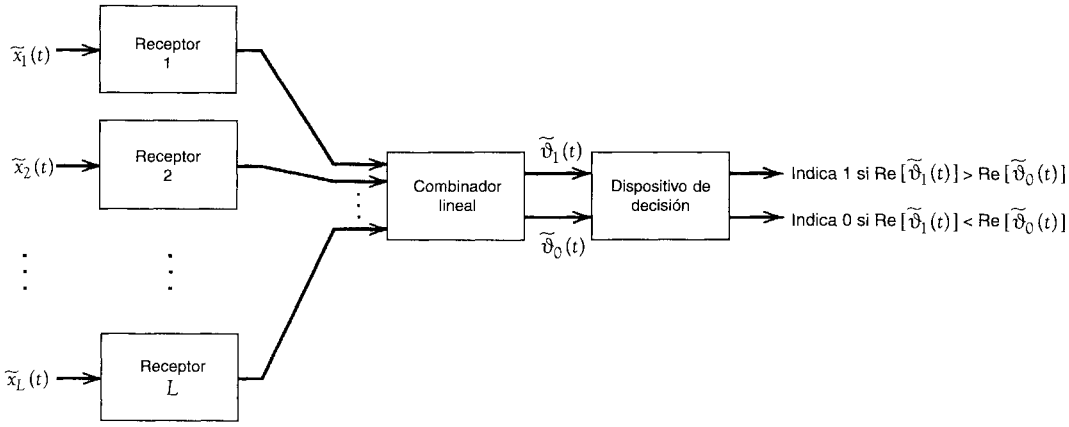


FIGURA 8.23 Diagrama de bloques que ilustra la técnica de diversidad del espacio.

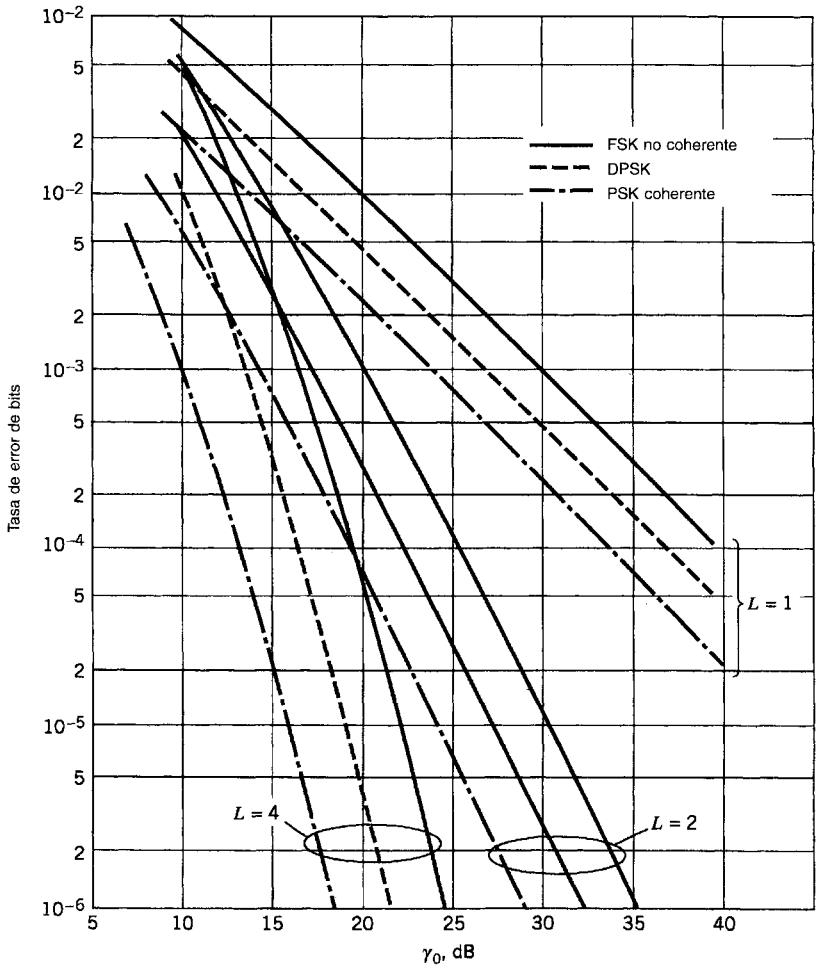


FIGURA 8.24 Desempeño de esquemas de transmisión de señales binarias con diversidad. (Tomado de Proakis con autorización de McGraw-Hill.)

ño óptimo en la transmisión de datos binarios (explicado aquí con fines ilustrativos) procediendo de la siguiente manera: la salida del k-ésimo filtro de acoplamiento en el ℓ-ésimo receptor, $v_{\ell k}(t)$, se multiplica por $\alpha_\ell \exp(j\phi_\ell)$, que representa a la conjugada compleja de la ℓ-ésima ganancia del canal, donde $\ell = 1, 2, \ldots, L$ y $k = 0, 1$. De ese modo, el combinador lineal produce dos envolventes complejas de salida definidas por

$$\tilde{v}_k(t) = \sum_{\ell=1}^{L} \alpha_\ell \exp(j\phi_\ell)\tilde{v}_{\ell k}(t), \qquad k = 0, 1 \tag{8.70}$$

de acuerdo con lo cual, $\alpha_\ell \exp(j\phi_\ell)$ desempeña la función de un *factor de ponderación*. Una envolvente compleja de salida $\tilde{v}_0(t)$ corresponde a la transmisión del símbolo 0, y la otra $\tilde{v}_1(t)$ corresponde a la transmisión del símbolo 1. Las partes reales de $\tilde{v}_0(t)$ y $\tilde{v}_1(t)$ se usan entonces en el proceso de toma de decisiones. Una situación que se describe aquí se aplica a la FSK binaria. En el caso de la PSK binaria, sólo se necesita un filtro de acoplamiento, situación en la cual un combinador lineal produce una envolvente compleja de salida sencilla. También en este caso, la parte real de la salida del combinador se utiliza en el proceso de toma de decisiones.

En el combinador lineal que se describe aquí, la relación de señal de salida instantánea a ruido (SNR) es la suma de las SNR instantáneas con respecto a las ramas (canales) de diversidad individuales. Esta forma óptima de combinador lineal es consecuentemente conocida como un *combinador de relación máxima*; vea el problema 8.17.

La figura 8.24 muestra el desempeño ante el ruido de la PSK binaria coherente, la DPSK binaria y la FSK binaria no coherente para $L = 2, 4$ canales de desvanecimiento independientes. Por el bien de la comparación hemos incluido en esta figura las gráficas correspondientes a un canal de desvanecimiento sin diversidad (es decir, $L = 1$). La figura 8.24 ilustra claramente la eficacia de la diversidad como un medio de mitigar los efectos a corto plazo del desvanecimiento de Rayleigh.

8.8 *Sistemas de comunicación inalámbricos TDMA y CDMA*[8]

En las comunicaciones inalámbricas, al igual que en la telefonía ordinaria, a un usuario le gustaría hablar y escuchar simultáneamente. Para satisfacer este deseo natural se requiere alguna forma de duplexado. Una forma en la cual es posible cumplir este requerimiento radica en proporcionar dos bandas de frecuencia, una para el enlace en sentido directo desde la estación base hasta una móvil y la otra para el enlace en sentido inverso desde el móvil hasta la estación base. Como se señaló antes, en Estados Unidos la sub-banda de 869–894 MHz se utiliza para el enlace en sentido directo, y la de 824–849 MHz, para el enlace inverso. Esta forma de duplexado recibe el nombre de *duplexado por división de frecuencia* (FDD). En realidad, el FDD es una parte integral de dos sistemas de comunicaciones inalámbricas ampliamente utilizados que se resumen en la tabla 8.3.

El primero de estos sistemas, a saber, el GSM, utiliza el acceso múltiple por división de tiempo (TDMA). De la sección 8.2 recordamos que en un sistema TDMA se permite a cada suscriptor tener acceso al canal de radio durante un conjunto de franjas de tiempo predeterminadas durante las cuales ese suscriptor particular tendrá el uso completo del canal. En consecuencia, los datos se transmiten por el canal en *ráfagas*, como se muestra en la *estructura de trama* de la figura 8.25. La trama básica del GSM se compone de ocho franjas de 577 μs. La *bandera* de 1 bit que precede a cada ráfaga de datos de 57 bits se usa para identificar si los bits de datos son habla digitalizada o alguna otra señal que contiene información. Los 3 bits de *cola*, todos ceros lógicos, se usan en la decodificación convolucional de los bits de datos codificados del canal. (Los códigos convolucionales se presentan en el capítulo 10.) La secuencia de entrenamiento de 26 bits a la mitad de la franja de tiempo se utiliza para la ecualización del canal. Por último, el *tiempo de seguridad*, que ocupa 8.25 bits, se incluye al final de cada franja para evitar

TABLA 8.3 ***Resumen de dos sistemas de comunicación inalámbricos utilizados ampliamente***

Aspecto	GSM*	*IS-95*†	*Comentarios*
Número de canales del dúplex	125	20	El CDMA asume 12.5 MHz en cada dirección; vea la siguiente línea
Ancho de banda del canal (kHz)	200	1250	
Tipo de acceso múltiple	TDMA	CDMA	
Usuarios del acceso por canal	8	20 a 35	Un sistema TDMA es *determinístico* en el sentido de que el número de usuarios del acceso por canal se define por medio del número de franjas de tiempo disponibles. Por otra parte, un sistema CDMA es de interferencia limitada en cuanto a que tiene un límite *flexible* en el número de usuarios del acceso por canal.
Tipo de modulación	GMSK	BPSK/QPSK	En el CDMA, los datos están modulados como BPSK, aunque la dispersión es QPSK
Velocidad de datos (kb/s)	270.833	9.6 o 14.4	
Periodo de la trama (ms)	4.615	20	En el CDMA, el periodo de la trama es igual al codec (codificador/decodificador) del habla

*GSM son las siglas en inglés de Sistema Global para Comunicaciones Móviles; originalmente se introdujo como un acrónimo de *Groupe de travail Spéciale pour les services Mobiles.*
†IS quiere decir *Interim Standard.*

que la ráfaga de datos recibida en la estación base proveniente de los móviles se traslape con algún otro; esto se consigue al no transmitir en absoluto señales durante el tiempo de seguridad. Con cada franja consistente en 156.25 bits, de los cuales 40.25 bits se pasan por alto (ignorando los 2 bits de bandera), la *eficiencia de trama* del GSM es

$$\left(1-\frac{40.25}{156.25}\right)\times 100 = 74.24\%$$

El segundo sistema de comunicación inalámbrico, IS-95, se resume en la tabla 8.3 y utiliza el acceso múltiple por división de código. De la sección 8.2 recordamos que en el CDMA, a cada suscriptor se le asigna un código de dispersión (secuencia de pseudo ruido) distinto, lo que permite el acceso total del suscriptor al canal durante todo el tiempo. En consecuencia, en un sistema CDMA tenemos una nueva forma de interferencia denominada *interferencia de acceso múltiple* (MAI), la cual se produce debido a la desviación de los códigos de dispersión con respecto a la ortogonalidad perfecta. Un fenómeno relacionado que requiere atención es el *problema de acercamiento-alejamiento*, que ocurre si las señales recibidas provenientes de las unidades móviles no tienen igual potencia en la

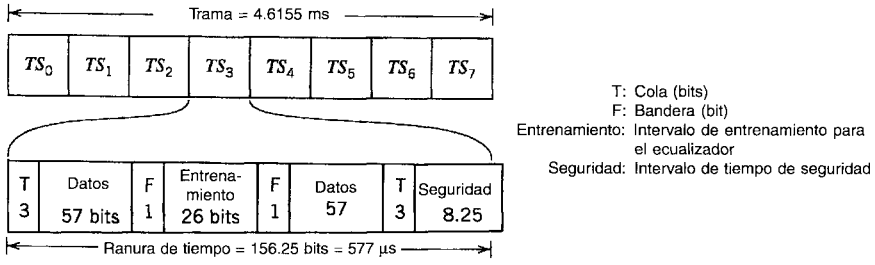


FIGURA 8.25 Estructura de trama del sistema de comunicación inalámbrico GSM.

estación base. En una situación de este tipo, la señal recibida más intensa de un usuario móvil captura el proceso de demodulación en la estación base en detrimento de otros usuarios. Para superar el problema de acercamiento-alejamiento, es usual utilizar *control de potencia* en la estación base, mediante el cual esta última mantiene el control sobre el nivel de potencia de la señal transmitida desde cualquier móvil que esté recibiendo el servicio de esa estación base. El uso de control de potencia es particularmente importante en los sistemas CDMA por otra razón. Un objetivo de los sistemas de acceso múltiple es maximizar la *capacidad del sistema*, misma que se define como el número más grande posible de usuarios que el sistema tiene la posibilidad de servir de manera confiable, dados recursos preestablecidos. Claramente se compromete la capacidad del sistema si cada móvil tiene libertad de aumentar su nivel de potencia transmitida sin importar los otros usuarios, ya que el aumento en la potencia transmitida elevará, a su vez, el nivel de interferencia de acceso múltiple en el sistema. Por tanto, para maximizar la capacidad del sistema es esencial que cada transmisor móvil se encuentre bajo el control de la estación base que lo atiende, de modo que la relación de señal a interferencia se mantenga en un nivel mínimo aceptable necesario para el servicio confiable.

Receptor RAKE

Resultaría incompleta una explicación de las comunicaciones inalámbricas utilizando el CDMA sin una descripción del *receptor RAKE*.[9] Este receptor se creó originalmente en la década de 1950 como un receptor de "diversidad" diseñado expresamente para ecualizar el efecto de la multitrayectoria. Lo primero, y más importante, es reconocer que la información útil en torno a la señal transmitida está contenida en el componente de multitrayectoria de la señal que se recibe. Por tanto, al considerar el punto de vista de que la multitrayectoria puede aproximarse como una combinación lineal de ecos retardados de manera diferente, el receptor RAKE busca combatir el efecto de la multitrayectoria utilizando un método de correlación para detectar de manera individual las señales de eco y sumarlas después algebraicamente. De este modo, la interferencia de intersímbolo debida a la multitrayectoria se encara reinsertando diferentes retardos en los ecos detectados, de manera que éstos realicen una función constructiva y no destructiva.

La figura 8.26 muestra la idea básica detrás del receptor RAKE. El receptor consiste en un número de *correlatores* conectados en paralelo y operando de manera síncrona. Cada correlator cuenta con dos entradas: 1) una versión retardada de la señal recibida y 2) una réplica de la

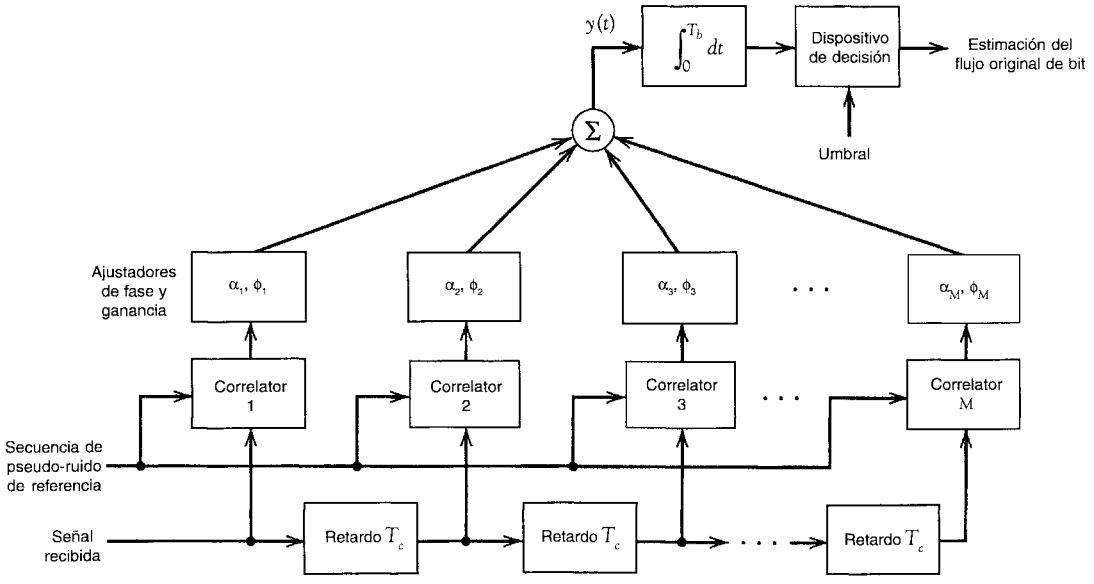


Figura 8.26 Diagrama de bloques del receptor RAKE.

secuencia de pseudo ruido que se utiliza como el código de dispersión para generar la señal modulada del espectro disperso en el transmisor. En efecto, la secuencia de pseudo ruido actúa como una "señal de referencia". Denotemos el ancho de banda nominal de la secuencia de pseudo ruido como $W = 1/T_c$, donde T_c es la duración de brizna. De la explicación de la modulación del espectro disperso que se presentó en el capítulo 7, recordamos que la función de autocorrelación de una secuencia de pseudo ruido tiene un pico único de ancho $1/W$, y que éste tiende hacia cero en cualquier lugar dentro de un periodo de la secuencia de pseudo ruido (es decir, un periodo del símbolo). De ese modo, sólo necesitamos hacer lo suficientemente grande el ancho de banda W de la secuencia de pseudo ruido para "identificar" los ecos significativos en la señal recibida. Para asegurar que la totalidad de las salidas del correlator se suman en forma constructiva, se efectúan otras dos operaciones en el receptor mediante bloques funcionales marcados como "ajustadores" de fase y ganancia:

1. Se introduce un retardo apropiado en cada salida de correlator, de modo que los ángulos de fase de las salidas del correlator concuerden entre sí.
2. Las salidas de correlator se ponderan de modo que los correlatores que responden a trayectorias intensas en el ambiente de multitrayectoria tengan sus contribuciones acentuadas, en tanto que los correlatores que no estén sincronizados con ninguna trayectoria importante se suprimen de modo correspondiente.

Los coeficientes de ponderación, α_k, se calculan de acuerdo con el *principio de combinación de relación máxima:*[10]

> La relación señal a ruido de una suma ponderada, donde cada elemento de la suma consiste en una señal más el ruido aditivo de potencia fija, se maximiza cuando la amplitud de ponderación se realiza en proporción a la intensidad de señal pertinente.

La salida del combinador lineal es

$$y(t) = \sum_{k=1}^{M} \alpha_k z_k(t) \tag{8.71}$$

donde $z_k(t)$ es la salida compensada de fase del k-ésimo correlator y M es el número de correlatores en el receptor. Siempre y cuando utilicemos suficientes correlatores en el receptor para expandir una región de retardos suficientemente ancha para abarcar todos los ecos significativos que probablemente ocurrirán en un ambiente de multitrayectoria, la salida $y(t)$ se comporta esencialmente como si hubiera una sola trayectoria de propagación entre el transmisor y el receptor en lugar de una serie de trayectorias múltiples dispersas en el tiempo.

Para simplificar la presentación, el receptor de la figura 8.26 supone el uso de corrimiento de fase por llaveo binaria al efectuar la modulación por espectro disperso en el transmisor. Por tanto, la operación final que se efectúa en la figura 8.26 es integrar la salida del combinador lineal $y(t)$ sobre el intervalo de bit T_b y determinar después si el símbolo binario 1 o 0 se transmitió en ese intervalo de bit.

El receptor RAKE deriva su nombre de que el banco de correlatores paralelos tiene una apariencia similar a los dedos de un rastrillo (rake). En virtud de que la modulación por espectro disperso resulta básica para la operación de comunicaciones inalámbrica CDMA, es natural que el receptor RAKE sea fundamental en el diseño del receptor que se utiliza en este tipo de comunicación de radio multiusuario.[11]

8.9 *Codificación de la fuente del habla para las comunicaciones inalámbricas*

Para el uso eficiente del ancho de banda del canal, los sistemas de comunicación inalámbrica digitales, sean del tipo TDMA o CDMA, se fundamentan en el uso de la *codificación del habla* para

eliminar casi toda la redundancia natural cuando se produce la voz, mientras se mantiene una alta calidad del habla en la decodificación. El método común implica utilizar codificación fuente, la cual, en una u otra forma, explota la *codificación predictiva lineal* (LPC) del habla.

En esta sección describimos dos técnicas diferentes para la codificación del habla: LPC excitada por pulsos múltiples y LPC excitada por código, cuyas versiones se usan respectivamente en el GSM y en el IS-95. Nuestro tratamiento de estas dos técnicas de codificación del habla se realiza en términos conceptuales.[12]

LPC EXCITADA POR PULSOS MÚLTIPLES

Esta forma de codificación del habla explota el *principio de análisis por síntesis*, el cual quiere decir que el codificador incluye una réplica del decodificador en su diseño. Específicamente, el codificador consiste en tres partes principales, como se indica en la figura 8.27*a*:

1. *Filtro de síntesis* para el modelado predictivo del habla. Éste puede incluir un filtro exclusivo de polo (es decir, un filtro cuya función de transferencia tiene únicamente polos), el cual se diseña para modelar la envolvente espectral *a corto plazo* del habla; el término *corto plazo* se refiere a que los parámetros del filtro se calculan con base en la predicción de la muestra presente de la señal del habla utilizando ocho de dieciséis muestras anteriores. El filtro de síntesis puede incluir también un predictor de *largo plazo* para modelar la estructura fina del espectro del habla; en tal caso, el predictor de largo plazo se conecta en cascada con el de corto plazo. En cualquier situación, la función del filtro de síntesis es producir una versión sintética del habla original que es de alta calidad.
2. *Generador de excitación* para producir la excitación que se aplica al filtro de síntesis. Esta última incluye un número definido de pulsos de cada 5 a 15 ms. Las amplitudes y las posiciones de los pulsos individuales son ajustables.
3. *Minimización del error* para optimizar el error ponderado perceptualmente entre el habla original y el habla sintetizada. El objetivo de esta minimización es optimizar las amplitudes y posiciones de los pulsos que se utilizan en la excitación. Por lo común se recurre a un criterio de error cuadrático medio para la minimización.

Por tanto, como se muestra en la figura 8.27*a*, las tres partes del codificador forman un procedimiento de optimización de *lazo cerrado*, que permite que el codificador opere a una velocidad de bit menor a 16 kb/s, en tanto se mantiene un habla de alta calidad.

El propio procedimiento de codificación incluye dos etapas principales:

- Los parámetros libres del filtro de síntesis se calculan utilizando las muestras reales del habla como entrada. Este cálculo se realiza fuera del lazo de optimización sobre un periodo de 10 a 30 ms, durante el cual la señal del habla se trata como pseudoestacionaria.

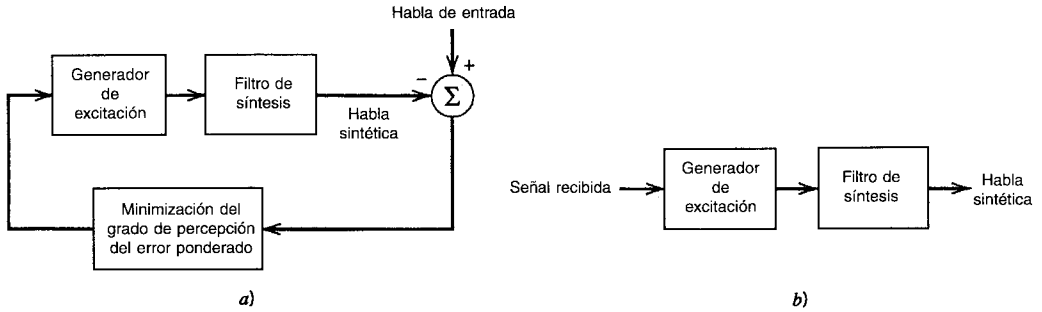


FIGURA 8.27 Codec predictivo lineal excitado por multipulso. *a*) Codificador. *b*) Decodificador (la señal recibida) cuya entrada consiste en los parámetros del filtro cuantizado y en la excitación cuantizada de acuerdo a como las produce el codificador.

- La excitación óptima para el filtro de síntesis se calcula minimizando el error ponderado perceptualmente con el lazo cerrado, como en la figura 8.27*a*.

De ese modo las muestras de voz se dividen en *tramas* (10 a 30 ms de largo) para calcular los parámetros del filtro, y cada trama se divide en forma adicional en *subtramas* (5 a 15 ms) para optimizar la excitación. Los parámetros del filtro cuantizado y de la excitación cuantizada constituyen la señal transmitida.

Advierta que al permitir primero que los parámetros del filtro varíen de una trama a la siguiente, y al permitir luego que la excitación varíe de una subtrama a la siguiente, se permite al codificador seguir el comportamiento no estacionario del habla, aunque en un esquema de lote por lote.

El decodificador, ubicado en el receptor, consiste simplemente en dos partes: generador de excitación y filtro de síntesis, como se ilustra en la figura 8.27*b*. Estas dos partes son idénticas a las correspondientes en el codificador. La función del decodificador es utilizar la señal recibida para producir una versión sintética de la señal del habla original. Lo anterior se consigue al pasar la excitación decodificada por el filtro de síntesis cuyos parámetros se igualan a los del codificador.

Al producir la complejidad computacional del codec (es decir, contracción de codificador/decodificador), los intervalos entre los pulsos individuales en la excitación se restringen para que asuman un valor común. El codec de análisis por síntesis que se produce se dice que tiene una *excitación de pulsos uniforme*.

■ LPC EXCITADO POR CÓDIGO

La figura 8.28 muestra el diagrama de bloques de un *LPC excitado por código*, conocido comúnmente como CELP. El rasgo distintivo del CELP es el uso de un *libro de códigos* predeterminado de vectores estocásticos (gaussianos blancos de media cero) como la fuente de excitación para el filtro de síntesis, el cual está compuesto por todos los filtros exclusivos de polo conectados en cascada; uno de ellos realiza la predicción a corto plazo y el otro la de largo plazo.

Como en el caso del LPC excitado por pulsos múltiples, los parámetros libres del filtro de síntesis se calculan primero, utilizando muestras reales del habla como entrada. A continuación, la elección de un vector (código) particular almacenado en el libro de códigos excitados y el factor de ganancia G en la figura 8.28 se optimizan minimizando la potencia promedio de error ponderado

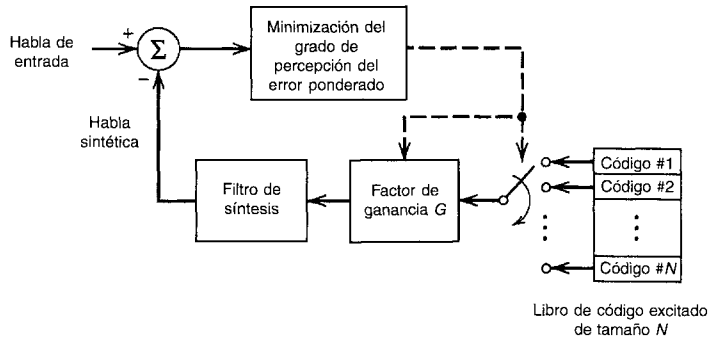


FIGURA 8.28 Codificador del codec predictivo lineal excitado por código (CELP): la señal transmitida consiste en el direccionamiento del código elegido del libro de códigos, la G cuantizada y los parámetros del filtro cuantizado.

perceptualmente entre el habla original y el habla sintetizada (es decir, la salida del filtro de síntesis). La dirección del vector estocástico elegido del libro de códigos y el factor de ganancia cuantizada correspondiente, junto con los parámetros cuantizados del filtro, constituyen la señal transmitida.

Una copia idéntica del libro de códigos se pone a disposición del decodificador, e igual ocurre para el filtro de síntesis. Por tanto, dada la señal recibida, se permite que el decodificador parametrice su propio filtro de síntesis y determine la excitación apropiada para este mismo, produciendo de esa manera una versión sintética de la versión del habla original.

La CELP es capaz de producir habla de buena calidad a velocidades de bit por abajo de 8 kb/s. Sin embargo, la complejidad computacional es considerable debido a la búsqueda exhaustiva del libro de códigos excitados. En particular, el habla sintetizada ponderada en el codificador tiene que calcularse para todas las entradas en el libro de códigos y luego compararse con el habla original ponderada. A pesar de eso, la puesta en práctica en tiempo real de los codec CELP se ha hecho posible en virtud de los avances en el procesamiento de señales digitales y en la tecnología VLSI.

8.10 *Arreglos de antenas adaptables para las comunicaciones inalámbricas*[13]

El objetivo de las comunicaciones inalámbricas es permitir a tantos usuarios como sea posible comunicarse en forma confiable independientemente de la ubicación y la movilidad. De la explicación que se presentó en las secciones 8.5 y 8.6, encontramos que esta meta se ve seriamente impedida por tres deterioros fundamentales del canal:

1. La *multitrayectoria* puede provocar un desvanecimiento severo debido a la cancelación de fase entre diferentes trayectorias de propagación. El desvanecimiento conduce a una reducción de la potencia disponible de la señal y, por tanto, a un desempeño degradado ante el ruido.
2. La *dispersión del retardo* resulta de las diferencias en los retardos de propagación entre las múltiples trayectorias de propagación. Cuando la dispersión del retardo excede aproximadamente 10% de la duración del símbolo, la interferencia de intersímbolo que experimenta la señal recibida alcanza un nivel importante, provocando una reducción en la velocidad de datos que puede obtenerse.
3. La *interferencia de cocanal* surge en los sistemas celulares donde los canales de frecuencia disponibles se dividen en diferentes conjuntos, cada uno asignándose a una celda específica y con varias celdas en el sistema utilizando el mismo grupo de frecuencias. La interferencia de cocanal limita la *capacidad del sistema* (es decir, el número mayor posible de usuarios que pueden ser atendidos confiablemente por el sistema).

Por lo común, los sistemas celulares utilizan sectorización de 120° en cada estación base, y sólo un usuario tiene acceso a un sector de una estación base a una frecuencia determinada. Podemos combatir los efectos del desvanecimiento de multitrayectoria y de la interferencia de cocanal en la estación base utilizando tres *arreglos de antena* idénticos aunque independientes, uno para cada sección de la estación base. La compensación de la dispersión del retardo se considera después en esta sección. La figura 8.29 muestra el diagrama de bloques de un *procesador de señales del arreglo*, donde se supone que hay N usuarios cuyas señales se reciben en un sector particular de la estación base, y el arreglo para ese sector consiste en M elementos de antena idénticos. Un usuario particular se trata como el único de interés, y los N-1 usuarios restantes originan la interferencia de cocanal. Además de esta última, cada componente de la entrada del procesador de señales del arreglo se corrompe por causa del ruido blanco gaussiano aditivo (AWGN). El análisis que se presenta aquí corresponde a las señales de la banda base, que en general son de valor complejo, lo que a su vez quiere decir que tanto el canal como el procesador de señales del arreglo requieren en cada caso caracterizaciones complejas. La estructura que se

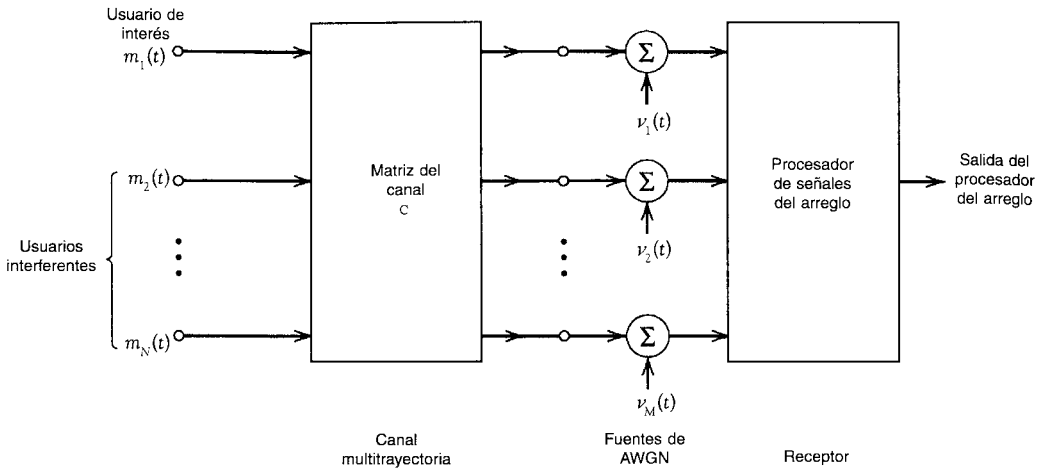


FIGURA 8.29 Diagrama de bloques del procesador de señales del arreglo que incluye M elementos de antena, y que está siendo activada por un canal de multitrayectoria.

describe en la figura 8.29 se dibuja para una salida relativa al usuario de interés. Este procesador de señales del arreglo se duplica para usuarios en otras frecuencias en la estación base.

El canal de multitrayectoria está caracterizado por la matriz del canal, la cual se denota mediante **C**. La matriz **C** tiene dimensiones M por N y, por tanto, puede expandirse en N vectores columna, como se indica mediante

$$\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N] \tag{8.72}$$

donde cada vector columna es de dimensión M.

Dada la configuración que se describe en la figura 8.29, la meta es diseñar un *procesador lineal de señales del arreglo* para el receptor, el cual satisfaga dos requerimientos:

1. Se cancela la interferencia de cocanal producida por los N-1 usuarios que interfieren.
2. La relación señal a ruido de salida (SNR) para el usuario de interés se hace máxima.

De aquí en adelante, estos dos requerimientos se citarán como los requerimientos de diseño 1 y 2.

Para proseguir con esta tarea de diseño, se supone que el canal multitrayectoria se describe mediante desvanecimiento de Rayleigh plano. En ese caso, a la luz del material que se presentó en la sección 8.7, encontramos que el uso de diversidad permite el tratamiento de los vectores columna $\mathbf{c}_1$, $\mathbf{c}_2$, ..., $\mathbf{c}_N$ como *linealmente independientes*, lo cual se justifica siempre y cuando el espaciamiento entre los elementos de la antena del arreglo sea lo suficientemente grande (por ejemplo siete veces la longitud de onda) para desvanecimiento independiente. Para simplificar la presentación, suponemos que el usuario 1 es el de interés y que los restantes $N - 1$ son responsables de la interferencia de cocanal, como se indica en la figura 8.29. El aspecto de diseño clave es cómo determinar el *vector de peso* denotado por **w**, el cual caracteriza al procesador de señales del arreglo. Para ese fin, procedemos de la manera siguiente:

1. Elegimos un vector de peso M–dimensional **w** que sea ortogonal a los vectores $\mathbf{c}_2$, ..., $\mathbf{c}_N$, los cuales se asocian a los usuarios que interfieren. Esta elección cumple con el requerimiento de diseño 1 (es decir, la cancelación de la interferencia de cocanal).
2. Para satisfacer el requerimiento de diseño 2 (es decir, la maximización de la SNR), nos apartamos un poco del tema presente para introducir la noción de un subespacio. Dado

un *espacio vectorial* o solo *espacio*, formado por un conjunto de vectores linealmente independientes, un *subespacio* del espacio es un subconjunto que satisface dos condiciones:[14]

i) Si sumamos cualesquiera dos vectores $\mathbf{z}_1$ y $\mathbf{z}_2$ en el subespacio, su suma, $\mathbf{z}_1$ y $\mathbf{z}_2$, sigue estando en el subespacio.

ii) Si multiplicamos cualquier vector $\mathbf{z}$ en el subespacio por cualquier escalar a, el múltiplo $a\mathbf{z}$ sigue estando en el subespacio.

Volviendo al asunto de cómo maximizar la SNR de salida para el usuario 1, construimos primero un subespacio denotado por $\mathcal{W}$, cuya dimensión es igual a la diferencia entre el número de elementos de la antena y el número de usuarios interferentes, esto es, $M - (N - 1) = M - N + 1$. A continuación, proyectamos el conjugado complejo del vector $\mathbf{c}_1$ del canal (perteneciente al usuario 1) sobre el subespacio $\mathcal{W}$. La proyección calculada de esa manera define el vector de peso $\mathbf{w}$.

Ejemplo 8.3

Para ilustrar el método del subespacio de dos etapas en la determinación del vector de peso $\mathbf{w}$, considere el ejemplo simple de un sistema que incluye dos usuarios caracterizados por los vectores de canal $\mathbf{c}_1$ y $\mathbf{c}_2$, y un arreglo de antena consistente en tres elementos; esto es, $N = 2$ y $M = 3$. En ese caso, para este ejemplo, el subespacio $\mathcal{W}$ es bidimensional, como indica

$$M - N + 1 = 3 - 2 + 1 = 2$$

Con el usuario 1 considerado como el de interés y el usuario 2 visto como el interferente, es posible construir el diagrama del espacio de señales como se muestra en la figura 8.30. El subespacio $\mathcal{W}$, indicado con el sombreado en esta figura, es ortogonal al vector $\mathbf{c}_2$ del canal. El vector de peso $\mathbf{w}$ del procesador de señales del arreglo se determina mediante la proyección del vector conjugado complejo del canal del usuario 1, esto es, $\mathbf{c}_1^*$, sobre el subespacio $\mathcal{W}$, como se describe en la figura 8.30. ◀

Una conclusión importante que se extrae de esta exposición es que un receptor lineal que utiliza la combinación óptima con M elementos de antena y que incluye $N - 1$ usuarios interferentes tiene el mismo desempeño que un receptor lineal con $M - N + 1$ elementos de antena sin interferencia, independientemente del ambiente de multitrayectoria. Para que pueda ponerse en prác-

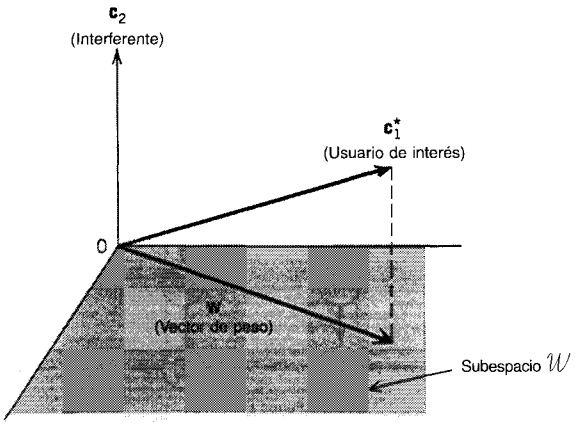


Figura 8.30 Diagrama del espacio de señales para el ejemplo 8.3, incluyendo un usuario de interés, un solo interferente y un arreglo de antena de tres elementos. El subespacio $\mathcal{W}$, que se muestra sombreado, es bidimensional en este ejemplo.

tica esta equivalencia, necesitamos desde luego que $M > N - 1$. Siempre y cuando se satisfaga esta condición, el receptor cancela la interferencia de cocanal con un mejoramiento de la diversidad igual a $M - N + 1$, lo que representa un incremento de N veces la capacidad del sistema.

El diseño de un procesador de señales del arreglo de acuerdo con el procedimiento de subespacio de dos etapas descrito aquí es del tipo de *forzamiento a cero*. Afirmamos lo anterior debido a que, dados M elementos de antena, el arreglo tiene suficientes grados de libertad para *forzar* la salida debido a $N - 1$ usuarios interferentes representados mediante los vectores del canal $\mathbf{c}_2, \ldots, \mathbf{c}_M$ linealmente independientes a cero siempre y cuando M sea mayor que $N - 1$. Advierta también que este procedimiento incluye $N = 1$ (es decir, un solo usuario sin usuarios interferentes) como un caso especial. En esta situación, la matriz del canal está compuesta por el vector $\mathbf{c}_1$, el cual yace en el subespacio $\mathcal{W}$, y la solución de forzamiento a cero $\mathbf{w}$ es igual a $\mathbf{c}_1^*$.

El análisis presentado hasta ahora ha sido completamente de tipo *espacial*, el cual ignora el efecto de la distribución del retardo. ¿Qué ocurre si la distribución del retardo es importante comparada con la duración del símbolo y por ello no puede ignorarse? Reconociendo que la distribución del retardo es responsable de la interferencia de intersímbolo, tenemos la posibilidad, a la luz del material que se presenta en el capítulo 4 sobre la ecualización de un canal telefónico, de incorporar un ecualizador *lineal* en cada rama de antena del arreglo para compensar la dispersión del retardo. El procesador de señales del arreglo que se produce toma la forma que se indica en la figura 8.31, la cual combina el procesamiento temporal y el espacial.

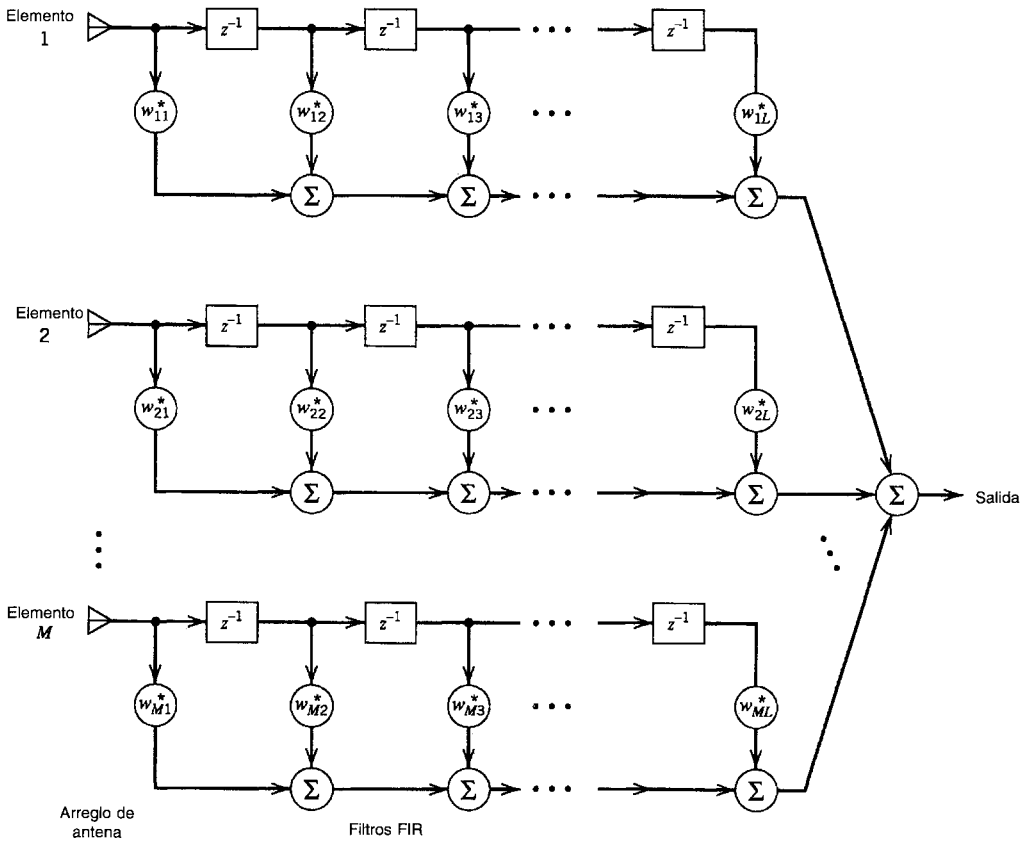


FIGURA 8.31 Procesador espacio-tiempo de la banda base. Los bloques denominados z^{-1} son elementos de retardo unitario, siendo cada retardo igual al periodo del símbolo. Los coeficientes del filtro son de valor complejo. Todos los filtros FIR se suponen de longitud L.

El procesamiento espacial lo proporciona el arreglo de la antena y el procesamiento temporal lo ofrece un banco de filtros de respuesta al impulso de duración finita (FIR). Por razones obvias, esta estructura recibe el nombre de *procesador espacio-tiempo*.[15]

■ ARREGLO DE ANTENA ADAPTABLE

El procedimiento de diseño del subespacio para el procesador de señales del arreglo en la figura 8.29 supone que los deterioros del canal son estacionarios y que conocemos la matriz **C** del canal. En realidad, el desvanecimiento de multitrayectoria, la dispersión del retardo y la interferencia del cocanal son todos, y cada uno de forma particular, no estacionarios. Además, es posible que la caracterización del canal se desconozca. Para encarar estos aspectos prácticos, necesitamos hacer adaptable el procesador de señales del arreglo de recepción de la figura 8.29. Considerando el alcance de este libro, delimitamos la exposición al procesamiento espacial adaptable, suponiendo que la dispersión del retardo puede ignorarse. Asimismo, suponemos que el fenómeno de desvanecimiento de multitrayectoria es lo suficientemente lento para justificar el *algoritmo de mínimos cuadrados* para llevar a cabo la adaptación.

La figura 8.32 muestra la estructura de un *arreglo de antena adaptable*, donde la salida de cada elemento de la antena se multiplica por un peso ajustable (controlable), y después las salidas de los elementos ponderados del arreglo se suman para producir la señal de salida del mismo. El arreglo de antena adaptable no requiere conocimiento de la dirección de arribo de las señales deseadas que se originan de un usuario de interés siempre y cuando el sistema se alimente con una *señal de referencia*, que está *correlacionada* con la señal deseada. La señal de salida del arreglo se sustrae de la señal de referencia para generar una *señal de error*, la cual se usa para aplicar los ajustes apropiados a los pesos elementales del arreglo. De este modo se incorpora en la operación del arreglo de la antena un sistema de realimentación para controlar los pesos elementales, lo que posibilita que ella se adapte a los cambios en el ambiente. Advierta que el diagrama de bloques de la figura 8.32 se dibuja para el procesamiento de la banda base, de lo que resulta la conjugación compleja de los pesos elementales. En un sistema práctico se usa cuadratura híbrida en cada elemento de la antena del arreglo para dividir la señal recibida de valor complejo en cada elemento en dos componentes: uno real y otro imaginario. El uso del híbrido se ha omitido en la figura 8.32 para simplificar el diagrama.

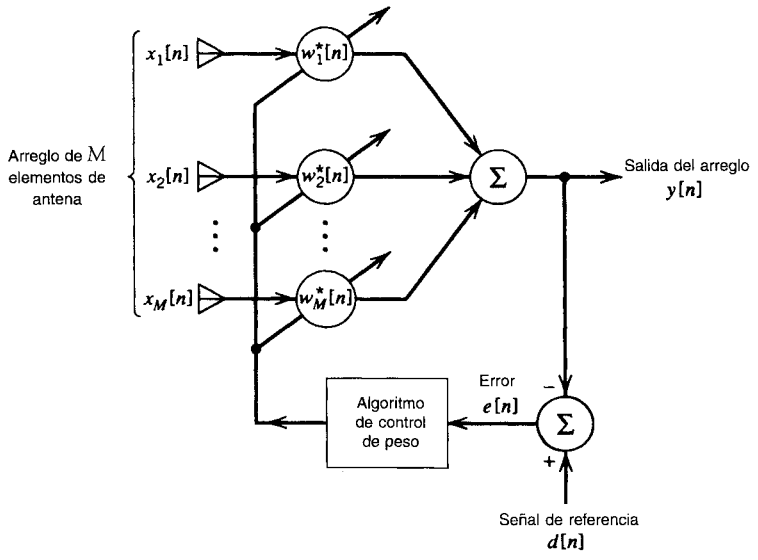


FIGURA 8.32 Diagrama de bloques del arreglo de antena adaptable.

Para optimizar el desempeño del arreglo de antena adaptable, es usual utilizar el *error cuadrático medio*

$$J = E\left[|e[n]|^2\right] \tag{8.73}$$

como la función de costo que debe *minimizarse*. La $e[n]$ es la señal de error en el tiempo $t = nT$, donde T es el periodo del símbolo y n es un entero que sirve como tiempo discreto. La minimización de la función de costo J suprime las señales de interferencia y mejora la señal deseada en la salida del arreglo. Sin embargo, el algoritmo de mínimos cuadrados minimiza el valor instantáneo de la función de costo J y, a través de iteraciones sucesivas, procura alcanzar el error cuadrático medio mínimo (mmse) (es decir, la solución óptima para los pesos elementales). De acuerdo con la explicación que se presentó en el capítulo 4 acerca de los ecualizadores temporales, los cuales llevan al dominio espacial, es posible afirmar que un arreglo de antena adaptable basado en el criterio del error cuadrático medio mínimo es altamente probable que proporcione una mejor solución que el basado en el criterio de forzamiento a cero incorporado en el método del subespacio de dos etapas.

Dejemos que $x_k[n]$ denote la salida del k-ésimo elemento en el arreglo al tiempo discreto n, y que $w_k[n]$ denote el valor correspondiente del peso conectado a este elemento. La señal de salida del arreglo (consistente en M elementos de antena) es entonces

$$y[n] = \sum_{k=1}^{M} w_k^*[n]x_k[n] \tag{8.74}$$

donde $w_k^*[n]x_k[n]$ es el producto interior de las cantidades de valor complejo $w_k[n]$ y $x_k[n]$. Al denotar la señal de referencia como $d[n]$, es posible evaluar la señal del error como

$$e[n] = d[n] - y[n] \tag{8.75}$$

Por tanto, el ajuste aplicado a cada k-ésimo peso elemental es

$$\Delta w_k[n] = \mu e^*[n]x_k[n], \quad k = 1, 2, \ldots, M \tag{8.76}$$

donde μ es el *parámetro del tamaño del escalón*, y el valor actualizado de este peso es

$$w_k[n+1] = w_k[n] + \Delta w_k[n], \quad k = 1, 2, \ldots, M \tag{8.77}$$

Las ecuaciones (8.74)-(8.77), en ese orden, constituyen el *algoritmo de mínimos cuadrados complejo*, que incluyen al algoritmo de mínimos cuadrados para señales reales (estudiado en los capítulos 3 y 4) como un caso especial. El algoritmo se inicia fijando $w_k[0] = 0$ para toda k. La derivación del algoritmo de mínimos cuadrados complejo se deja como el problema 8.19.

Son tres las ventajas del arreglo de antena adaptable utilizando el algoritmo de mínimos cuadrados:

- Simplicidad en la puesta en práctica.
- Crecimiento lineal en la complejidad con el número de elementos de antena.
- Desempeño robusto con respecto a las perturbaciones.

Sin embargo, el sistema sufre de las siguientes desventajas:

- Velocidad de convergencia baja, la cual corresponde por lo común a diez veces el número de pesos. Esto limita el uso del algoritmo de mínimos cuadrados complejo a un ambiente de desvanecimiento lento, para el cual la dispersión Doppler es pequeña en comparación con el recíproco de la duración del intervalo de observación.
- La sensibilidad del comportamiento de la convergencia ante las variaciones de las potencias de la señal de referencia y de la interferencia de cocanal.

Estas limitaciones del algoritmo de mínimos cuadrados complejo pueden superarse utilizando un algoritmo conocido como *inversión de matrices directa*, el cual surge directamente del filtro de Wiener, que se explica en el capítulo 4; vea el problema 8.21. A diferencia del algoritmo

de mínimos cuadrados, el de inversión de matrices directa opera en el modo de *lote* en el sentido de que el cálculo de los pesos elementales se basa en un lote de K instantáneas. El tamaño del lote K se elige como un compromiso entre dos requerimientos en conflicto:

- El tamaño K debe ser suficientemente pequeño para que el lote de instantáneas utilizadas en el cálculo sea tratado de modo justificable como pseudoestacionario.
- El tamaño K debe ser lo suficientemente grande para los valores calculados de los pesos elementales para aproximarse a la solución del error cuadrático medio mínimo.

El algoritmo de inversión directa de matriz es la técnica de combinación óptima para antenas de arreglo que se despliegan en muchas estaciones base en la actualidad. Dicho algoritmo puede reformularse para el cálculo recursivo,[16] si se desea.

Cuando el teletráfico es elevado, las estaciones de base se configuran ordinariamente como microceldas, las cuales son celdas pequeñas, tales como un piso de oficina o una estación desplegada a lo largo de una autopista con antenas direccionales. En una configuración de este tipo, hay muchas estaciones de base económicas muy próximas una de otra. El uso de arreglos de antena adaptables proporciona los medios para una configuración alternativa en la que hay menor número (aunque más costosas) de estaciones base y más apartadas entre sí que en el sistema microcelular correspondiente.

8.11 *Resumen y discusión*

En este capítulo explicamos dos importantes tipos de comunicaciones multiusuario: de satélite e inalámbricas. Los sistemas de comunicación por satélite ofrecen cobertura mundial, en tanto que los inalámbricos proporcionan movilidad. La cobertura mundial y la movilidad que dan estos dos sistemas de comunicación han transformado de manera profunda la manera en la que nos comunicamos, tanto local como mundialmente.

Aunque los sistemas de comunicación por satélite e inalámbricos funcionan de maneras por completo diferentes, ambos se fundamentan en la propagación de radio para enlazar al receptor con el transmisor. En las comunicaciones por satélite tenemos un *enlace ascendente* desde una terminal terrena hasta el transpondedor de satélite y un *enlace descendente* desde el satélite hasta otra terminal terrena. El satélite opera como un repetidor en el cielo. Además, con el satélite ubicado en una órbita geoestacionaria, los enlaces ascendente y descendente operan como trayectorias de línea de visión de longitudes fijas. Por tanto, el canal de comunicación por satélite, abarcando ambos enlaces, se modela con mucha exactitud como un *canal de ruido blanco gaussiano aditivo (AWGN)*.

El sistema de comunicación inalámbrica tiene también dos enlaces propios: uno *ascendente* o *enlace inverso* para la transmisión del móvil a la estación base, uno *descendente* o *enlace directo* para la transmisión de la estación base al móvil. La estación base está fija, localizándose en el centro o en el borde de una región de cobertura. Consiste en canales de radio y en antenas de transmisor y receptor montadas sobre una torre. Las tres principales fuentes de degradación en las comunicaciones inalámbricas, explicadas en el capítulo, son la interferencia de cocanal, el desvanecimiento y la dispersión del retardo; las dos últimas son subproductos de la multitrayectoria. Una característica común de estos deterioros del canal es que todos son *fenómenos dependientes de la señal*. A diferencia del ubicuo ruido del canal, los efectos degradantes de interferencia y multitrayectoria no pueden combatirse consecuentemente al aumentar simplemente la potencia de la señal transmitida. En vez de eso, ambos requieren del uso de técnicas especializadas, hechas a la medida de sus características físicas particulares. Estas técnicas especializadas incluyen la diversidad, las antenas de arreglo adaptable y el receptor RAKE.

Cerramos la exposición con señalamientos que contrastan la comunicación inalámbrica con la comunicación alámbrica. Del capítulo 3 recordamos que la principal fuente de interés en los

sistemas de la comunicación alámbrica es el ruido; estos sistemas tienen el suficiente ancho de banda de canal para permitir el uso de modulación por codificación de pulsos (PCM) como el método estándar para convertir el habla en un flujo de 64 kb/s, lo cual proporciona los datos básicos para un desempeño casi sin ruido. En cambio, en la comunicación inalámbrica, el ancho de banda del canal constituye un valioso recurso, cuya conservación implica el uso de técnicas de codificación del habla espectralmente eficientes para producir habla digitalizada de calidad de conversación a velocidades que son una fracción pequeña de la correspondiente a la PCM. Infortunadamente, los codificadores de forma de onda ejemplificados para la modulación por codificación de pulsos diferencial, que se explicó en el capítulo 3, no satisfacen este estricto requerimiento. El método que se prefiere radica en utilizar técnicas de codificación de fuente espectralmente eficiente: codificación predictiva lineal (LPC) excitada por pulsos múltiples o su variante excitada por pulsos uniformes y la LPC excitada por código (CELP); estas técnicas de codificación de fuente producen velocidades de bit inferiores a 16 kb/s para eliminar casi toda la redundancia natural en el habla, lo cual mantiene una alta calidad de la misma aunque de tipo sintético. Para proporcionar protección contra el ruido se recurre a la codificación del canal mediante la cual se insertan bits redundantes en el flujo de datos transmitidos de una manera controlada. El uso de la codificación del canal también ayuda de otras maneras: extiende el intervalo de aparatos de baja potencia, así como la vida de las baterías. La codificación de canal se explica en el capítulo 10.

NOTAS Y BIBLIOGRAFÍA

1. En lo que respecta a un tratamiento detallado de las comunicaciones por satélite y de aspectos relacionados, consulte los siguientes libros: Sklar [140], Pratt y Bostian [122], Wu [173], Bhargava y otros [18] y Spilker, Jr. [145]. El primero, tercero, cuarto y quinto libros subrayan el uso de satélites en las comunicaciones digitales. El libro de Pratt y Bostian presenta un amplio tratamiento de las comunicaciones por satélite, remarcando aspectos tan diversos como la propagación de ondas de radio, las antenas, la mecánica orbital, el procesamiento de señales y la electrónica de radio.
2. El análisis del presupuesto del enlace se expone en los libros de Sklar [140] y Anderson [7]; en cuanto a las comunicaciones por satélite éste se explica en Bhargava y otros [18].
3. En cuanto a los fundamentos de las antenas, vea el libro de Kraus [80] y el capítulo 11 del libro de Jordan y Balmain.
4. La ecuación del espacio libre (8.14) se conoce así gracias a Friis. Para el origen de la fórmula de Friis de la ecuación (8.30), vea Friis [252].
5. En lo referente a un tratamiento original del radio celular, vea el artículo de MacDonald [315].
6. En el libro de Parsons [115] se presenta un tratamiento amplio del canal de propagación de radio móvil. Este libro presenta los fundamentos de la propagación VHF y UHF, la propagación sobre terrenos irregulares y áreas con construcciones, así como una caracterización estadística del canal de radio móvil. Este último tema se presenta también en Proakis [123], donde se ofrece una explicación legible del efecto del desvanecimiento sobre el desempeño del error de los canales de desvanecimiento de Rayleigh y una buena explicación de las técnicas de diversidad. En lo que se refiere a un tratamiento completo del tema, consulte los capítulos 9-11 de Stein en el libro editado por Schwartz, Bennett y Stein [133].
7. La distribución ji cuadrada es un caso especial de la distribución gamma. La función de densidad de probabilidad de una *variable aleatoria con distribución gamma de* X tiene dos parámetros: $\alpha > 0$ y $\lambda > 0$; ésta se define mediante

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad 0 < x < \infty$$

donde $\Gamma(\alpha)$ es la *función gamma*, la cual se define por medio de

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-\alpha} dz, \quad \alpha > 0$$

La función gamma tiene las siguientes propiedades:

$$\Gamma(1/2) = \sqrt{\pi}$$
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0$$

Dejando $\lambda = 1/2$ y $\alpha = k/2$, donde k es un entero positivo, obtenemos la *distribución ji cuadrada* con $2k$ grados de libertad, como muestra

$$f_X(x) = \frac{x^{(k-2)/2} e^{-x/2}}{2^{k/2}\Gamma(k/2)}, \quad 0 < x < \infty$$

8. Para un artículo general acerca de la evolución de las comunicaciones inalámbricas, vea Oliphant [334]. En lo que se refiere a libros sobre los fundamentos de los sistemas de comunicación inalámbricos, consulte Steele y Hanzo [149], Stüber y Rappaport [127].

   En lo que se refiere a una descripción detallada del GSM, vea el capítulo 8 del libro de Steele y Hanzo [149]. En lo relativo a una descripción detallada del sistema IS-95, vea el libro de Lee y Miller [86].

9. El artículo clásico acerca del receptor RAKE, lo elaboraron Price y Green [344].

10. En relación con un artículo original acerca de cómo maximizar la relación de señal a ruido en forma práctica a partir de la suma de varias señales ruidosas, vea el artículo clásico de Brennan.

11. La aplicación del receptor RAKE en los sistemas de comunicación inalámbricos CDMA se analiza en detalle en el libro de Viterbi.

12. La idea de la excitación de pulsos múltiples para la codificación del habla se debe a Atal y Remde [190]. La predicción lineal excitada por código (CELP) del habla fue introducida por primera vez por Atal y Schroeder [191]. Para un análisis matemático detallado de los tipos de codificación del habla de excitación por pulsos múltiples, de excitación por pulsos uniformes y de excitación por código, relacionadas en particular con las comunicaciones inalámbricas, vea el capítulo 3 del libro editado por Steele y Hanzo [149].

13. En la literatura de las comunicaciones inalámbricas, los arreglos de antenas adaptables a menudo se refieren como *antenas inteligentes*. En cuanto a un panorama de los diversos aspectos implicados en el uso de los arreglos de antena adaptables para comunicaciones inalámbricas, vea el artículo y las notas de curso de Winters [405, 406]. El procedimiento del subespacio de dos etapas para diseñar el procesador de señales del arreglo de la figura 8.29 se basa en el material que se presenta en Winters [406]; el libro de Rappaport (1999) presenta una colección de artículos acerca de los arreglos de antena adaptables, los cuales se agrupan en algoritmos, arquitecturas, aplicaciones de hardware, modelos de canal y evaluación del desempeño.

14. La idea del subespacio está implantada en el álgebra de matrices. En cuanto a un análisis de esta idea, consulte a Strang y Stewart. En lo que se refiere a una explicación de la descomposición del subespacio en el contexto del procesamiento estadístico de señales, refiérase a Scharf.

15. En lo referente a explicaciones tutelares del procesamiento de espacio-tiempo en las comunicaciones inalámbricas, vea los artículos de Paulraj y Ng [340], Paulraj y Papadias [341] y Kohno [297].

16. La puesta en práctica recursiva del algoritmo de la inversión directa de matrices condujo a un nuevo algoritmo conocido comúnmente como el *algoritmo de mínimos cuadrados recursivo (RLS)*; en cuanto a la deducción de dicho algoritmo y sus variantes vea Haykin [62].

PROBLEMAS

Propagación por el espacio libre

8.1 Un enlace de radio utiliza un par de antenas de disco de 2m con una eficiencia cada una de 60%, como antenas de transmisión y recepción. Otras especificaciones del enlace son:

Potencia transmitida = 1 dBW
Frecuencia de portadora = 4 GHz
Distancia del receptor
desde el transmisor = 150 m

Calcule *a*) la pérdida en el espacio libre, *b*) la ganancia de potencia de cada antena y *c*) la potencia recibida en dBW.

8.2 Repita el problema 8.1 para una frecuencia de portadora de 12 GHz.

8.3 La ecuación (8.14) es una formulación de la ecuación de Friis del espacio libre. Demuestre que esta ecuación también puede formularse en las siguientes formas equivalentes:

a) $P_r = \dfrac{P_t A_t A_r}{\lambda^2 d^2}$

b) $P_r = \dfrac{P_t A_t G_r}{4\pi d^2}$

donde P_t es la potencia transmitida, A_t es el área efectiva de la antena de transmisión, λ es la longitud de onda de la portadora, d es la distancia del receptor a partir del transmisor, G_r es la ganancia de potencia de la antena de recepción, A_r es el área efectiva de la antena de recepción y P_r es la potencia recibida.

Analice las situaciones que favorecen el uso de una de estas ecuaciones sobre las otras.

8.4 A partir de la definición matemática de la pérdida en el espacio libre

$$L_{\text{espacio libre}} = \left(\frac{4\pi d}{\lambda}\right)^2$$

vemos que ésta es dependiente de la longitud de onda λ o la frecuencia f de la portadora. ¿Cómo puede justificarse en términos físicos la dependencia de la longitud de onda o la frecuencia?

8.5 En un sistema de comunicaciones por satélite, la frecuencia de la portadora que se usa en el enlace ascendente siempre es superior que la del enlace descendente. Justifique las razones de esta elección.

8.6 Un transmisor de radio faro de onda continua se localiza sobre un satélite en órbita geoestacionaria. La salida de 12 GHz del radio faro se supervisa por medio de una estación terrestre ubicada a 40,000 km del satélite. La antena de transmisión de este último es un disco de 1 m con una eficiencia de apertura de 70%, y la antena receptora de la estación terrestre es un disco de 10 m con una eficiencia de apertura de 55%. Calcule la potencia recibida, puesto que la potencia de salida del radio faro corresponde a 100 mW.

Figura de ruido

8.7 Considere un resistor de 75 Ω que se mantiene a "temperatura ambiente" de 290K. Suponiendo un ancho de banda de 1 MHz calcule lo siguiente:

a) El valor cuadrático medio del voltaje que aparece en las terminales de este resistor debido al ruido térmico.

b) La máxima potencia de ruido disponible que se entrega a una carga acoplada.

8.8 En este problema volvemos al ejemplo 8.1 con base en la configuración de receptor de la figura 8.10. Suponga que una guía de onda con pérdida se inserta entre la antena receptora y el amplificador de ruido bajo. La pérdida de la guía de onda corresponde a 1 dB y su temperatura física es igual a 290K. Recalcule la temperatura de ruido efectiva del receptor.

8.9 Considere el receptor de la figura P8.9, el cual está compuesto con una guía de onda con pérdida, amplificador RF de ruido bajo, convertidor descendente de frecuencia (mezclador) y amplificador FI. La ilustración incluye las figuras de ruido y las ganancias de potencia de estos cuatro componentes. La temperatura de la antena es de 50K.

a) Calcule la temperatura de ruido equivalente en cada uno de los cuatro componentes de la figura P8.9, suponiendo una temperatura ambiente $T = 290$K.

b) Calcule la temperatura de ruido efectiva del receptor completo.

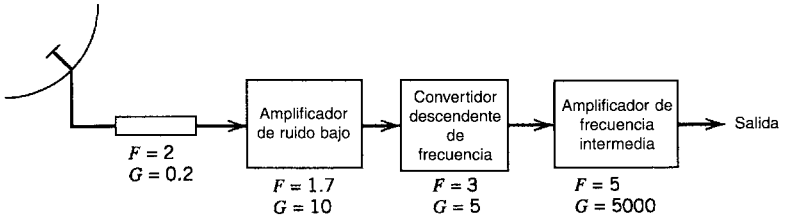


FIGURA P8.9

Cálculos del presupuesto del enlace

8.10 En este problema nos enfocamos en el presupuesto de la potencia del enlace ascendente del sistema de comunicación digital por satélite considerado en el ejemplo 8.2. Los parámetros del enlace son los siguientes:

Frecuencia de portadora	= 14 GHz
Densidad de potencia en el amplificador TWT en la saturación	= –81 dBW/m^2
Figura de mérito del satélite, G/T	= 1.9 dB/K
Distancia del satélite desde la terminal terrestre transmisora	= 40,000 km

a) Suponiendo que no hay respaldo de potencia del TWT, calcule la proporción C/N_0 en el satélite.

b) Dado que la velocidad de los datos en el enlace ascendente es la misma que la calculada para el enlace descendente en el ejemplo 8.2, estime la probabilidad del error de símbolo en el que se incurre en el enlace ascendente al permitir un margen del enlace de 6 dB. Compare su resultado con el del ejemplo 8.2.

8.11 La proporción C/N_0 del enlace descendente en un sistema de satélite de transmisión directa (DBS) se estima igual a 85 dB-Hz. Las especificaciones del enlace son:

EIRP del satélite	= 57 dBW
Frecuencia de la portadora del enlace descendente	= 12.5 GHz
Velocidad de datos	= 10 Mb/s
E_b/N_0 requerido en la terminal terrestre receptora	= 10 dB
Distancia del satélite a partir de la terminal terrestre receptora	= 41,000 km

Calcule el diámetro mínimo de la antena de disco necesaria para proporcionar una recepción de TV satisfactoria, suponiendo que el disco tiene una eficiencia de 55% y está localizado junto a una casa donde la temperatura es de 310K. En este cálculo, suponga que la operación del sistema DBS está limitada esencialmente al enlace descendente.

Comunicaciones inalámbricas

8.12 Tanto las comunicaciones inalámbricas como las de satélite se fundamentan en la propagación de las ondas de radio para sus operaciones. Resuma *a)* las similitudes de estos dos sistemas de comunicación multiusuario, y *b)* las principales diferencias que distinguen a uno del otro.

8.13 En los sistemas de comunicación inalámbrica, la frecuencia de portadora del enlace ascendente (enlace inverso) es más pequeña que la frecuencia de portadora del enlace descendente (enlace directo). Justifique las razones de esta diferencia.

8.14 La figura P8.14 describe las trayectorias directa (línea de visión) e indirecta (reflejada) de un enlace de radio que opera sobre un plano terrestre. Las alturas de la antena transmisora en la estación base y la antena receptora en la unidad móvil son, respectivamente, h_b y h_m. Suponga lo siguiente:

- El coeficiente de reflexión del suelo es -1.
- La distancia d entre las dos antenas es suficientemente grande para hacer lo suficientemente pequeña la diferencia de fase ϕ entre las trayectorias reflejada y directa en comparación con un radián, de manera que podemos fijar sen $\phi \simeq \phi$.

Por tanto, demuestre que la potencia recibida P_r está dada por la aproximación

$$P_r \simeq P_t G_b G_m \left(\frac{h_b h_m}{d^2}\right)^2$$

donde P_t es la potencia transmitida y G_b y G_m son las ganancias de potencia de las antenas de transmisión de la base y móvil, respectivamente. Compare este resultado con la ecuación de Friis del espacio libre.

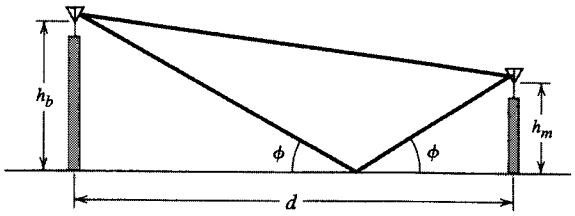


FIGURA P8.14

8.15 El *modelo de dos trayectorias* definido por la respuesta al impulso

$$h(t) = a_1\ \delta(t-\tau_1) + a_2\ \exp(-j\theta)\ \delta(t-\tau_2)$$

se usa frecuentemente en el tratamiento analítico de los sistemas de comunicación inalámbricos. Los parámetros del modelo son los tiempos de retardo τ_1 y τ_2, la fase θ distribuida uniformemente y los coeficientes reales a_1 y a_2.

a) Determine: *i)* la función de transferencia del modelo, *ii)* su perfil de retardo de potencia.

b) Demuestre que el modelo exhibe desvanecimiento selectivo de frecuencia debido a variaciones en los coeficientes a_1 y a_2.

8.16 En el receptor RAKE ilustrado en la figura 8.26, cada correlator se sincroniza insertando el retardo correcto en la señal recibida.

a) Demuestre que, en teoría, se obtiene el mismo resultado insertando el retardo correcto en la señal de referencia (es decir, la secuencia de pseudo ruido).

b) En la práctica, el método preferido consiste en utilizar el procedimiento que se describe en la figura 8.26. ¿Qué razón sugiere usted para esta preferencia?

8.17 En este problema estudiamos el esquema de diversidad de combinación de razón máxima. Para proceder, consideramos un conjunto de señales ruidosas $\{x_j(t)\}_{j=1}^{N}$, donde $x_j(t)$ se define por medio de

$$x_j(t) = s_j(t) + n_j(t), \qquad j = 1, 2, \ldots, N$$

Suponga lo siguiente:

- Los componentes de la señal $s_j(t)$ son coherentes localmente, esto es,

$$x_j(t) = \zeta_j m(t), \qquad j = 1, 2, \ldots, N$$

donde las ζ_j son números reales positivos y $m(t)$ denota una señal de mensaje con potencia unitaria.

- Las componentes de ruido $n_j(t)$ tienen media cero y son estadísticamente independientes, esto es,

$$E[n_j(t)n_k(t)] = \begin{cases} \sigma_j^2 & \text{para } k = j \\ 0 & \text{en otro caso} \end{cases}$$

La salida del combinador lineal está definida por

$$x_j(t) = \sum_{j=1}^{N} \alpha_j x_j(t)$$

donde los parámetros α_j van a determinarse.

a) Demuestre que la relación de señal de salida a ruido es

$$(\text{SNR})_O = \frac{\left(\sum_{j=1}^{N} \alpha_j x_j\right)^2}{\sum_{j=1}^{N} \alpha_j^2 \sigma_j^2}$$

b) Deje que

$$u_j = \alpha_j \sigma_j$$

$$v_j = \frac{\zeta_j}{\sigma_j}$$

y reformule la expresión para $(\text{SNR})_O$. Por tanto, aplicando la desigualdad de Schwarz a esta reformulación, demuestre que

i) $(\text{SNR})_O \leq \sum_{j=1}^{N} (\text{SNR})j$

donde $(\text{SNR})_j = \zeta_j^2 / \sigma_j^2$.

ii) los valores óptimos de los coeficientes del combinador se definen mediante

$$\alpha_j = \frac{\zeta_j}{\sigma_j^2}$$

en cuyo caso se satisface la desigualdad de Schwarz con el signo de igualdad.

La desigualdad de Schwarz se explica en la sección 5.2.

Arreglos de antena adaptables

8.18 Considere el procesador de señales del arreglo de la figura 8.29 donde sólo hay dos usuarios ($N = 2$) y el arreglo está compuesto por dos elementos ($M = 2$). Construya el subespacio $\mathcal{W}$ para este problema. Por tanto, utilizando un diagrama del espacio de señales, ilustre el cálculo del peso que caracteriza al procesador de señales del arreglo.

8.19 En este problema obtenemos el algoritmo de mínimos cuadrados complejo. Con referencia a la figura 8.32 y empezando con la función de costo instantáneo

$$J = \frac{1}{2}|e[n]|^2$$

donde $e[n]$ es la señal del error y M es el número de elementos de antena, realice lo siguiente:

a) Determine la derivada de la función de costo J con respecto al k-ésimo peso elemental $w_k[n]$.

b) Utilizando la derivada instantánea $\partial J/\partial w_k[n]$, denotada por $\hat{\nabla}J[k]$, determine el ajuste $\Delta w_k[n]$ hecho al k-ésimo peso elemental de acuerdo con la regla

$$\Delta w_k[n] = -\mu\hat{\nabla}J[k]$$

c) Verifique la composición del algoritmo de mínimos cuadrados complejo descrito en las ecuaciones de la (8.75) a la (8.77).

Advierta que $w_k[n]$ es de valor complejo, y que usted necesita considerar por separado sus partes real e imaginaria.

8.20 Una limitación práctica de un arreglo de antena adaptable utilizando el algoritmo de mínimos cuadrados es el intervalo dinámico sobre el cual puede operar el arreglo. Esta limitación se debe a que la velocidad de respuesta de los pesos en el algoritmo de mínimos cuadrados es proporcional a la potencia promedio de la señal a la entrada del arreglo.

a) Justifique la afirmación de que el intervalo dinámico de la potencia promedio de la señal a la entrada del arreglo es proporcional a $R_b/f_{máx}$, donde R_b es la velocidad de datos en b/s y $f_{máx}$ es la velocidad de desvanecimiento máxima en Hz.

b) Suponiendo un factor de proporcionalidad de 0.2, por medio del cual se escala la proporción $R_b/f_{máx}$, calcule el intervalo dinámico de un arreglo de antena adaptable utilizando el algoritmo de mínimos cuadrados para $R_b = 32$ kb/s y $f_{máx} = 70$ Hz. Comente su resultado. (El factor de proporcionalidad de 0.2 es una elección razonable para sistemas que utilizan PSK.)

8.21 En este problema obtenemos el algoritmo de inversión directa de matriz para ajustar los pesos del arreglo de antena adaptable. Para hacerlo de esa manera, regresamos a la deducción del filtro de Wiener que se presentó en el capítulo 3.

a) Demuestre que

$$\hat{\mathbf{R}}_x\mathbf{w} = \hat{\mathbf{r}}_{xd}$$

donde $\hat{\mathbf{R}}_x$ es una estimación de la matriz de correlación del vector de entrada $\mathbf{x}[k]$:

$$\hat{\mathbf{R}}_x\frac{1}{K} = \sum_{k=1}^{K}\mathbf{x}[k]\mathbf{x}^H[k]$$

y $\hat{\mathbf{r}}_{xd}$ es una estimación del vector de correlación cruzada entre $\mathbf{x}[k]$ y la señal de referencia $d[k]$:

$$\hat{\mathbf{r}}_{xd} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}[k]d^*[k]$$

El exponente H en la fórmula para $\hat{\mathbf{R}}_x$ denota la transportación hermetiana (es decir, transposición y conjugación compleja), de modo que $\mathbf{x}[k]\mathbf{x}^H[k]$ denota el producto exterior de $\mathbf{x}[k]$ consigo mismo. Las sumatorias tanto para $\hat{\mathbf{R}}_x$ como para $\hat{\mathbf{r}}_{xd}$ se llevan a cabo sobre un total de K instantáneas, con cada una de ellas representada por el par $\{\mathbf{x}[k], d[k]\}$.

b) Utilizando las fórmulas de la parte *a*, describa un algoritmo para calcular el vector de peso $\mathbf{w}$, dado un conjunto de datos compuesto de K instantáneas. Por tanto, demuestre que la complejidad del algoritmo crece como M^3 con el tamaño del vector de peso $\mathbf{w}$ denotado por M.

CAPÍTULO 9

LÍMITES FUNDAMENTALES EN LA TEORÍA DE LA INFORMACIÓN

El memorable artículo de Shannon sobre la teoría de la información en 1948, así como los refinamientos acerca del mismo efectuados por otros investigadores, constituyeron una respuesta directa a la necesidad de los ingenieros eléctricos de diseñar sistemas de comunicación que fueran tanto eficientes como confiables. La comunicación eficiente desde una fuente hasta el destino del usuario se consigue a través de la codificación de la fuente. La comunicación confiable por un canal ruidoso se alcanza por medio de la codificación de control de errores. Este capítulo se enfoca en estos importantes aspectos, como se resume a continuación:

- *La entropía como la medida básica de información.*
- *El teorema de codificación de fuente y los algoritmos de compactación de datos.*
- *Información mutua y su relación con la capacidad de un canal de comunicación para la transmisión de información.*
- *El teorema de comunicación del canal como la base para la comunicación confiable.*
- *El teorema de la capacidad de información como la base para el compromiso entre el ancho de banda del canal y la relación señal a ruido.*
- *La teoría de distorsión de la velocidad para la codificación de fuente con un criterio de fidelidad.*

9.1 *Introducción*

Como se mencionó en el capítulo de antecedentes y preliminares y se reiteró a lo largo del texto, el propósito de un sistema de comunicación es transmitir señales de banda base que contienen información desde un lugar a otro por un canal de comunicación. En los capítulos precedentes del libro hemos descrito una diversidad de esquemas de modulación para cumplir este objetivo. Sin embargo, ¿qué es lo que entendemos por el término *información*? Para considerar este aspecto necesitamos apelar a la *teoría de la información*.[1] Esta disciplina de amplia base matemática ha efectuado aportaciones fundamentales, no sólo a las comunicaciones, sino también a la ciencia del cómputo, la física estadística y la inferencia estadística, así como a la probabilidad y la estadística.

En el contexto de las comunicaciones, la teoría de la información tiene que ver con el modelado y el análisis matemático de un sistema de comunicación y no con los canales y las fuentes físicos. En particular, esto proporciona respuestas a dos preguntas fundamentales (entre otras):

- ¿Cuál es la complejidad irreductible debajo de la cual no es posible comprimir una señal?
- ¿Cuál es la velocidad de transmisión última para la comunicación confiable por un canal ruidoso?

Las respuestas a estas preguntas se encuentran en la entropía de una fuente y en la capacidad de un canal, respectivamente. La *entropía* se define en términos del comportamiento probabilístico de una fuente de información; se nombra de esa manera en deferencia al uso paralelo de este concepto en la termodinámica. La *capacidad* se define como la posibilidad intrínseca de un canal para transportar información; se relaciona de manera natural con las características de ruido del canal. Un resultado notable que surge de la teoría de la información es que si la entropía de la fuente es menor que la capacidad del canal, entonces puede conseguirse la comunicación sin errores por el canal. Por tanto, resulta conveniente iniciar nuestro estudio de la teoría de la información explicando las relaciones entre incertidumbre, información y entropía.

9.2 *Incertidumbre, información y entropía*

Suponga que un *experimento probabilístico* implica la observación de la salida emitida por una fuente discreta durante cada unidad de tiempo (intervalo de transmisión de señales). La salida de la fuente se modela como una variable aleatoria discreta, S, la cual toma los símbolos de un *alfabeto* finito fijo

$$\mathscr{S} = \{s_0, s_1, \ldots, s_{K-1}\} \tag{9.1}$$

con probabilidades

$$P(S = s_k) = p_k, \quad k = 0, 1, \ldots, K-1 \tag{9.2}$$

Desde luego, este conjunto de probabilidades debe satisfacer la condición

$$\sum_{k=0}^{K-1} p_k = 1 \tag{9.3}$$

Suponemos que los símbolos emitidos por la fuente durante intervalos de transmisión de señales sucesivas son estadísticamente independientes. Una fuente que tiene las propiedades que acaban de señalarse recibe el nombre de *fuente discreta sin memoria*; sin memoria porque el símbolo emitido en cualquier tiempo es independiente de las elecciones anteriores.

¿Es posible que encontremos una medida de cuánta información produce una fuente de este tipo? Para responder la pregunta, advertimos que la idea de información se relaciona estrechamente con la de incertidumbre o sorpresa, según se describe a continuación.

Considere el evento $S = s_k$, que describe la emisión del símbolo s_k por la fuente con probabilidad p_k, según se define en la ecuación (9.2). Claramente, si la probabilidad $p_k = 1$ y $p_i = 0$ para toda $i \neq k$, entonces no hay "sorpresas" y, por tanto, ninguna "información" cuando se emite el símbolo s_k, ya que sabemos cuál debe ser el mensaje de la fuente. Si, en cambio, los símbolos de la fuente ocurren con probabilidades diferentes y la probabilidad p_k es baja, entonces hay más sorpresa, y consecuentemente información, cuando la fuente emite el símbolo s_k que cuando emite el símbolo s_i, $i \neq k$, con probabilidad superior. Por tanto, las palabras *incertidumbre, sorpresa* e *información* están relacionadas. Antes de que ocurra el evento $S = s_k$, hay una cantidad de incertidumbre. Cuando ocurre el evento $S = s_k$, hay una cantidad de sorpresa. Después de la ocurrencia del evento $S = s_k$, existe una ganancia en la cantidad de información, cuya esencia puede considerarse como la *resolución de la incertidumbre*. Además, la cantidad de información se relaciona con el *inverso* de la probabilidad de ocurrencia.

Definimos la cantidad de información ganada después de observar el evento $S = s_k$, el cual ocurre con probabilidad p_k, como la función *logarítmica*[2]

$$I(s_k) = \log\left(\frac{1}{p_k}\right) \tag{9.4}$$

Esta definición exhibe las siguientes propiedades importantes que se satisfacen de manera intuitiva:

1.
$$I(s_k) = 0 \qquad \text{para } p_k = 1 \tag{9.5}$$
Evidentemente, si estamos absolutamente *seguros* del resultado de un evento, incluso antes de que ocurra, *no* se gana información.

2.
$$I(s_k) \geq 0 \qquad \text{para } 0 \leq p_k \leq 1 \tag{9.6}$$
Es decir, la ocurrencia de un evento $S = s_k$ proporciona o no información, pero nunca origina una *pérdida* de información.

3.
$$I(s_k) > I(s_i) \qquad \text{para } p_k < p_i \tag{9.7}$$
Es decir, cuanto menos probable es un evento, tanto más información se gana cuando éste ocurre.

4. $I(s_k s_l) = I(s_k) + I(s_l)$ si s_k y s_l son estadísticamente independientes.

La base del logaritmo en la ecuación (9.4) es bastante arbitraria. No obstante, es una práctica estándar actual utilizar un logaritmo de base 2. La unidad de información que resulta recibe el nombre de *bit* [una contracción del término en inglés *bi*nary digit (dígito binario)]. De tal modo escribimos

$$\begin{aligned} I(s_k) &= \log_2\left(\frac{1}{p_k}\right) \\ &= -\log_2 p_k \qquad \text{para } k = 0, 1, \ldots, K-1 \end{aligned} \tag{9.8}$$

Cuando $p_k = 1/2$, tenemos $I(s_k) = 1$ bit. Por tanto, *un bit es la cantidad de información que ganamos cuando ocurre uno de dos eventos posibles e igualmente probables.* Observe que la información $I(s_k)$ es positiva, ya que el logaritmo de un número menor que uno, tal como una probabilidad, es negativo.

La cantidad de información $I(s_k)$ producida por la fuente durante un intervalo de transmisión de señales arbitrario depende del símbolo s_k emitido por la fuente en ese tiempo. En realidad, $I(s_k)$ es una variable aleatoria discreta que toma los valores $I(s_0)$, $I(s_1)$,..., $I(s_{k-1})$ con probabilidades p_0, p_1,..., p_{k-1} respectivamente. La media de $I(s_k)$ sobre el alfabeto de la fuente $\mathscr{S}$ está dada por

$$\begin{aligned} H(\mathscr{S}) &= E[I(s_k)] \\ &= \sum_{k=0}^{K-1} p_k I(s_k) \\ &= \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \end{aligned} \tag{9.9}$$

La cantidad importante $H(\mathscr{S})$ recibe el nombre de *entropía*[3] de una fuente discreta sin memoria con alfabeto de fuente. Ésta es una medida del *contenido de información promedio por símbolo de la fuente.* Advierta que la entropía $H(\mathscr{S})$ depende sólo de las probabilidades del símbolo en el alfabeto $\mathscr{S}$ de la fuente. De ese modo, el símbolo $\mathscr{S}$ en $H(\mathscr{S})$ no es argumento de una función sino más bien una etiqueta para una fuente.

■ ALGUNAS PROPIEDADES DE LA ENTROPÍA

Considere una fuente discreta sin memoria cuyo modelo matemático está definido por las ecuaciones (9.1) y (9.2). La entropía $H(\mathscr{S})$ de una fuente de este tipo está acotada de la manera siguiente:

$$0 \leq H(\mathscr{S}) \leq \log_2 K \tag{9.10}$$

donde K es la *base* (número de símbolos) del alfabeto de la fuente. Además, es posible plantear dos enunciados:

1. $H(\mathscr{S}) = 0$, si y sólo si la probabilidad $p_k = 1$ para alguna k, y las probabilidades restantes en el conjunto son todas cero; esta cota inferior de la entropía corresponde a *ninguna incertidumbre.*
2. $H(\mathscr{S}) = \log_2 K$, si y sólo si $p_k = 1/K$ para toda k (es decir, todos los símbolos en el alfabeto $\mathscr{S}$ son *igualmente probables*); esta cota superior de la entropía corresponde a *incertidumbre máxima.*

Para probar estas propiedades de $H(\mathscr{S})$, procedemos de la siguiente forma. Primero, puesto que cada probabilidad p_k es menor que la unidad, se sigue que cada término $p_k \log_2(1/p_k)$ en la ecuación (9.9) siempre es no negativo, y por ello $H(\mathscr{S}) \geq 0$. Luego, advertimos que el término del producto $p_k \log_2(1/p_k)$ es cero si, y solamente si, $p_k = 0$ o 1. Por tanto, deducimos que $H(\mathscr{S}) = 0$, si y sólo si, $p_k = 0$ o 1, es decir, $p_k = 1$ para cierta k y todas las demás son cero.

Esto completa las demostraciones de la cota inferior en la ecuación (9.10) y el enunciado (1).

Para probar la cota superior en la ecuación (9.10) y el enunciado (2), recurrimos a la propiedad del logaritmo natural:

$$\log x \leq x - 1, \quad x \geq 0 \tag{9.11}$$

Esta desigualdad puede verificarse de inmediato graficando las funciones $\log x$ y $(x - 1)$ en función de x, como se ilustra en la figura 9.1. Aquí observamos que la línea $y = x - 1$ siempre se encuentra sobre la curva $y = \log x$. La igualdad se cumple *sólo* en el punto $x = 1$, donde la línea es tangente a la curva.

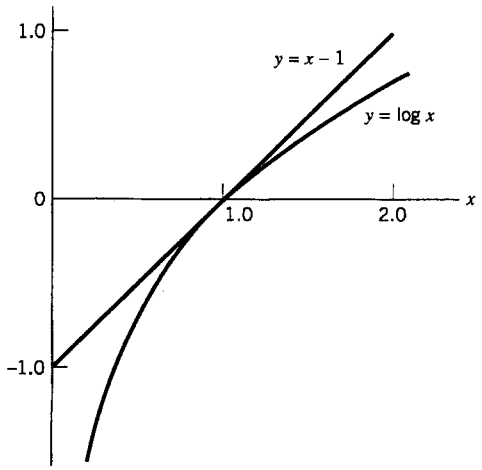


FIGURA 9.1 Gráficas de las funciones $x - 1$ y $\log x$ en función de x.

Para continuar con la demostración, considere primero cualesquiera dos distribuciones de probabilidad $\{p_0, p_1, ..., p_{K-1}\}$ y $(q_0, q_1, ..., q_{K-1}\}$ en el alfabeto $\mathscr{S} = \{s_0, s_1, ..., s_{K-1}\}$ de una fuente discreta sin memoria. En ese caso, cambiando al logaritmo natural, podemos escribir

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) = \frac{1}{\log 2}\sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right)$$

Por tanto, utilizando la desigualdad de la ecuación (9.11), obtenemos

$$\begin{aligned}\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) &\leq \frac{1}{\log 2}\sum_{k=0}^{K-1} p_k\left(\frac{q_k}{p_k} - 1\right)\\ &\leq \frac{1}{\log 2}\sum_{k=0}^{K-1}(q_k - p_k)\\ &\leq \frac{1}{\log 2}\left(\sum_{k=0}^{K-1} q_k - \sum_{k=0}^{K-1} p_k\right) = 0\end{aligned}$$

De ese modo tenemos una *desigualdad fundamental*

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) \leq 0 \tag{9.12}$$

donde la igualdad se cumple sólo si $q_k = p_k$ para toda k.

Suponga que a continuación dejamos

$$q_k = \frac{1}{K}, \qquad k = 0, 1, \ldots, K-1 \tag{9.13}$$

que corresponde a un alfabeto $\mathscr{S}$ con símbolos *igualmente probables*. La entropía de una fuente discreta sin memoria con una caracterización tal es igual a

$$\sum_{k=0}^{K-1} q_k \log_2\left(\frac{1}{q_k}\right) = \log_2 K \tag{9.14}$$

Además, el uso de la ecuación (9.13) en la (9.12) produce

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \leq \log_2 K$$

De manera equivalente, la entropía de una fuente discreta sin memoria con una distribución de probabilidad arbitraria para los símbolos de su alfabeto $\mathscr{S}$ está acotada como

$$H(\mathscr{S}) \leq \log_2 K$$

Por consiguiente, $H(\mathscr{S})$ siempre es menor o igual que $\log_2 K$. La igualdad se cumple sólo si los símbolos en el alfabeto $\mathscr{S}$ son igualmente probables, como en la ecuación (9.13). Esto completa la prueba de la ecuación (9.10) y los enunciados (1) y (2).

▶ EJEMPLO 9.1 Entropía de una fuente sin memoria binaria

Para ilustrar las propiedades de $H(\mathscr{S})$ consideramos una fuente binaria para la cual el símbolo 0 ocurre con probabilidad p_0 y el símbolo 1 con probabilidad $p_1 = 1 - p_0$. Suponemos que la fuente es sin memoria, de modo que los símbolos sucesivos emitidos por ella son independientes estadísticamente.

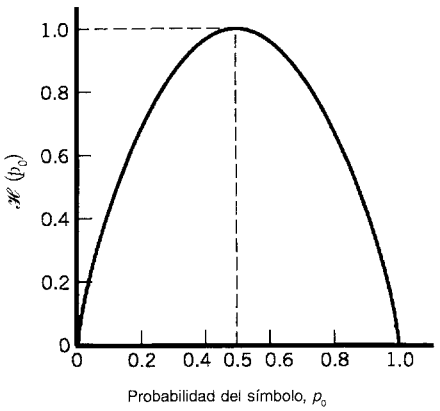


FIGURA 9.2 Función de entropía $\mathscr{H}(p_0)$.

La entropía de una fuente de estas características es igual a

$$\begin{aligned} H(\mathscr{S}) &= -p_0 \log_2 p_0 - p_1 \log_2 p_1 \\ &= -p_0 \log_2 p_0 - (1-p_0) \log_2 (1-p_0) \text{ bits} \end{aligned} \tag{9.15}$$

a partir de lo cual observamos lo siguiente:

1. Cuando $p_0 = 0$, la entropía $H(\mathscr{S}) = 0$; esto sigue del hecho de que $x \log x \to 0$ cuando $x \to 0$.
2. Cuando $p_0 = 1$, la entropía $H(\mathscr{S}) = 0$.
3. La entropía $H(\mathscr{S})$ alcanza su valor máximo, $H_{\text{máx}} = 1$ bit, cuando $p_1 = p_0 = 1/2$, es decir, los símbolos 1 y 0 son igualmente probables.

La función de p_0 dada en el segundo miembro de la ecuación (9.15) se encuentra con frecuencia en los problemas de la teoría de la información. Por tanto, suele asignársele un símbolo especial a esta función. Específicamente, definimos

$$\mathscr{H}(p_0) = -p_0 \log_2 p_0 - (1-p_0) \log_2 (1-p_0) \tag{9.16}$$

Nos referimos a $\mathscr{H}(p_0)$ como la *función de entropía*. La distinción entre la ecuación (9.15) y la (9.16) debe señalarse con todo cuidado. La $H(\mathscr{S})$ de la ecuación (9.15) produce la entropía de una fuente discreta sin memoria con alfabeto de fuente $\mathscr{S}$. La $\mathscr{H}(p_0)$ de la ecuación (9.16), en cambio, es una función de la probabilidad previa p_0 definida en el intervalo [0, 1]. Por tanto, es posible graficar la función entropía $\mathscr{H}(p_0)$ en función de p_0, definida en el intervalo [0, 1], como en la figura 9.2. La curva en esta última figura subraya las observaciones efectuadas de acuerdo con los puntos 1, 2 y 3. ◀

■ EXTENSIÓN DE UNA FUENTE DISCRETA SIN MEMORIA

Al discutir los conceptos teóricos de la información, encontramos a menudo útil considerar *bloques* en vez de símbolos individuales, con cada bloque consistiendo en n símbolos de fuentes sucesivos. Podemos considerar cada bloque de este tipo como si fuera producido por una *fuente extendida* con un alfabeto de fuente $\mathscr{S}^n$ que tiene K^n bloques *distintos*, donde K es el número de símbolos distintos en el alfabeto de fuente $\mathscr{S}$ de la fuente original. En el caso de una fuente discreta sin memoria, los símbolos de fuente son estadísticamente independientes. Por ello, la probabilidad de un símbolo de fuente en $\mathscr{S}^n$ es igual al producto de las probabilidades de los n símbolos de fuente $\mathscr{S}$ que constituyen el símbolo de fuente particular en $\mathscr{S}^n$. Es posible por ello

esperar intuitivamente que $H(\mathscr{S}^n)$, la entropía de la fuente extendida, sea igual a n veces $H(\mathscr{S})$, la entropía de la fuente original. Es decir, podemos escribir

$$H(\mathscr{S}^n) = nH(\mathscr{S}) \tag{9.17}$$

▶ EJEMPLO 9.2 Entropía de la fuente extendida

Considere una fuente discreta sin memoria con alfabeto de fuente $\mathscr{S} = \{s_0, s_1, s_2\}$ con probabilidades respectivas

$$p_0 = \tfrac{1}{4}$$
$$p_1 = \tfrac{1}{4}$$
$$p_2 = \tfrac{1}{2}$$

Por tanto, el uso de la ecuación (9.9) produce la entropía de la fuente como

$$\begin{aligned} H(\mathscr{S}) &= p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) \\ &= \frac{1}{4}\log_2(4) + \frac{1}{4}\log_2(4) + \frac{1}{2}\log_2(2) \\ &= \frac{3}{2}\text{bits} \end{aligned}$$

Considere a continuación la extensión de segundo orden de la fuente. Con el alfabeto de fuente $\mathscr{S}$ consistente en tres símbolos, se sigue que el alfabeto de fuente $\mathscr{S}^2$ de la fuente extendida tiene nueve símbolos. El primer renglón de la tabla 9.1 presenta los nueve símbolos de $\mathscr{S}^2$, denotados $\sigma_0, \sigma_1, \ldots, \sigma_8$. El segundo renglón de la tabla presenta la composición de estos nueve símbolos en términos de las secuencias correspondientes de los símbolos de fuente s_0, s_1 y s_2, tomados dos a la vez. Las probabilidades de los nueve símbolos de fuente de la fuente extendida se presentan en el último renglón de la tabla. Por tanto, el uso de la ecuación (9.9) produce la entropía de la fuente extendida como

$$\begin{aligned} H(\mathscr{S}^2) &= \sum_{i=0}^{8} p(\sigma_i) \log_2 \frac{1}{p(\sigma_i)} \\ &= \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) \\ &\quad + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{4}\log_2(4) \\ &= 3\text{ bits} \end{aligned}$$

De tal manera vemos que $H(\mathscr{S}^2) = 2H(\mathscr{S})$ de acuerdo con la ecuación (9.17).

TABLA 9.1 *Particularidades del alfabeto de la extensión de segundo orden de una fuente discreta sin memoria*

Símbolos de $\mathscr{S}^2$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Secuencias correspondientes de símbolos de $\mathscr{S}$	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Probabilidad de $i = 0, 1, \ldots, 8$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

9.3 *Teorema de codificación de fuente*

Un importante problema en las comunicaciones es la representación *eficiente* de los datos generados por una fuente discreta. El proceso mediante el cual se lleva a cabo esta representación recibe el nombre de *codificación de fuente*. El dispositivo que lleva a cabo la representación recibe el nombre de un *codificador de fuente*. Para que el codificador de fuente sea *eficiente* necesitamos conocimiento de la estadística de la fuente. En particular, si se sabe que algunos símbolos de la fuente son más probables que otros, entonces es posible explotar esta característica en la generación de un *código fuente* asignando palabras de código *cortas* a los símbolos de fuente *frecuentes*, y palabras de código *largas* a los símbolos de fuente *esporádicos*. Nos referimos a un código de fuente de este tipo como *código de longitud variable*. El *código Morse* constituye un ejemplo de un código de longitud variable. En este famoso código, las letras del alfabeto y los números se codifican en secuencias de *marcas* y *espacios*, denotados como puntos "." y guiones "–", respectivamente. En el lenguaje inglés, la letra *E* ocurre con mayor frecuencia que la letra *Q*, por ejemplo, por lo que el código Morse codifica a la *E* en un solo punto ".", la palabra de código más corta, y codifica a la *Q* en "– – . –", la palabra de código más larga.

Nuestro interés principal radica en el desarrollo de un codificador de fuente eficiente que satisfaga dos requerimientos funcionales:

1. Las palabras de código producidas por el codificador están en forma *binaria*.
2. El código fuente es *decodificable en forma única*, por lo que la secuencia de fuente original puede reconstruirse perfectamente a partir de la secuencia binaria codificada.

Considere entonces el esquema que se muestra en la figura 9.3, el cual describe una fuente discreta sin memoria cuya salida s_k es convertida por el codificador de fuente en un bloque de ceros y unos, denotado por b_k. Suponemos que la fuente tiene un alfabeto con K símbolos diferentes, y que el k-ésimo símbolo s_k ocurre con probabilidad p_k, $k = 0, 1, ..., K - 1$. Considere que la palabra de código binario asignada al símbolo s_k por el codificador tiene longitud l_k, medida en bits. Definimos la longitud promedio de palabra de código, $\bar{L}$, del codificador de fuente como

$$\bar{L} = \sum_{k=0}^{K-1} p_k l_k \tag{9.18}$$

En términos físicos, el parámetro $\bar{L}$ representa el *número promedio de bits por símbolo de fuente* utilizado en el proceso de codificación de fuente. Sea $L_{\text{mín}}$ el valor *mínimo* posible de $\bar{L}$. Definimos entonces la *eficiencia de codificación* del codificador de fuente como

$$\eta = \frac{L_{\text{mín}}}{\bar{L}} \tag{9.19}$$

Con $\bar{L} \geq L_{\text{mín}}$, tenemos claramente que $\eta \leq 1$. Se dice que el codificador de fuente será *eficiente* cuando η se acerque a la unidad.

Sin embargo, ¿cómo se determina el valor mínimo $L_{\text{mín}}$? La respuesta a esta pregunta fundamental está incorporada en el primer teorema de Shannon: el *teorema de la codificación de fuente*,[4] el cual puede enunciarse del modo siguiente:

> Dada una fuente discreta sin memoria de entropía $H(\mathscr{S})$, la longitud promedio de palabra de código $\bar{L}$ para cualquier esquema de codificación de fuente sin distorsión está acotada como
>
> $$\bar{L} \geq H(\mathscr{S}) \tag{9.20}$$

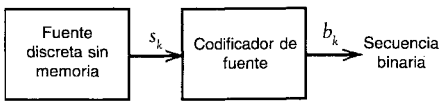


FIGURA 9.3 Codificación de puente.

(Una prueba de este teorema para una clase particular de códigos de fuente se presenta en la siguiente sección.) De acuerdo con el teorema de codificación de fuente, la entropía $H(\mathscr{S})$ representa un *límite fundamental* sobre el número promedio de bits por símbolo de fuente necesario para representar una fuente discreta sin memoria en el sentido de que puede hacerse tan pequeño, aunque no menor que, la entropía $H(\mathscr{S})$. De manera que con $L_{\text{mín}} = H(\mathscr{S})$, es posible reescribir la eficiencia de un codificador de fuente en términos de la entropía $H(\mathscr{S})$ como

$$\eta = \frac{H(\mathscr{S})}{\bar{L}} \tag{9.21}$$

9.4 Compactación de datos

Una característica común de las señales generadas por fuentes físicas es que, en su forma natural, contienen una cantidad importante de información que es *redundante*, y cuya transmisión es, por tanto, un desperdicio de los recursos de comunicación principales. En la transmisión *eficiente* de señales, *es necesario eliminar la información redundante de la señal antes de la operación.* Esta operación, *sin* pérdida de información, suele efectuarse sobre una señal en forma digital, en cuyo caso nos referiremos a ella como *compactación de datos* o *compresión de datos sin pérdida.* El código que resulta de una operación de este tipo proporciona una representación de la salida de la fuente que no es sólo eficiente en términos del número promedio de bits por símbolo, sino también es exacta porque los datos originales pueden reconstruirse sin pérdida de información. La entropía de la fuente establece el límite fundamental sobre la remoción de redundancia de los datos. Básicamente, la compactación de datos se consigue asignando descripciones cortas a los resultados más frecuentes de la salida de la fuente y descripciones largas a los menos frecuentes.

En esta sección explicamos algunos esquemas de codificación de fuente para compactación de datos. Empezamos la exposición describiendo un tipo de código de fuente conocido como código de prefijo, que no sólo es decodificable sino que también ofrece la posibilidad de producir una longitud promedio de palabra de código que puede hacerse arbitrariamente cercana a la entropía de la fuente.

Codificación de prefijo

Considere una fuente discreta sin memoria de alfabeto $\{s_0, s_1, ..., s_{K-1}\}$ y estadísticas $\{p_0, p_1, ..., p_{K-1}\}$. Para que un código fuente que represente la salida de esta fuente sea de uso práctico, dicho código tiene que ser decodificable en forma única. La restricción anterior asegura que para cada secuencia finita de símbolos emitida por la fuente, la secuencia correspondiente de palabras de código sea diferente de la secuencia de palabras de código correspondiente a cualquier otra secuencia de fuente. Estamos interesados de modo específico en una clase especial de códigos que satisfacen una restricción conocida como la *condición de prefijo.* Para definir la condición de prefijo, considere que la palabra de código asignada al símbolo fuente s_k se denota por medio de $(m_{k1}, m_{k2}, ..., m_{kn})$, donde los elementos individuales $m_{k1}, ..., m_{ki}$ son ceros y unos, y n es la longitud de la palabra de código. La parte inicial de la palabra de código se representa mediante los elementos $m_{k1}, ..., m_{ki}$ para cierta $i \leq n$. Cualquier secuencia conformada por la parte inicial de la palabra de código recibe el nombre de *prefijo* de la palabra de código. Un *código de prefijo* se define como el código en el cual ninguna palabra de código es el prefijo de cualquier otra palabra de código.

Para ilustrar el significado de un código de prefijo, considere los tres códigos fuente que se describen en la tabla 9.2. El código I no es de prefijo, pues el bit 0, la palabra de código para s_0, es un prefijo de 00, la palabra de código para s_2. De igual manera, el bit 1, la palabra de código correspondiente a s_1, es un prefijo de 11, la palabra de código de s_3. Similarmente, es posible demostrar que el código III no es un código de prefijo, aunque el código II lo sea.

Para decodificar una secuencia de palabras de código generadas a partir de un código fuente de prefijo, el *decodificador de fuente* se inicia simplemente al principio de la secuencia y decodifica

TABLA 9.2 *Ilustración de la definición de un código de prefijos*

Símbolo de la fuente	*Probabilidad de ocurrencia*	*Código I*	*Código II*	*Código III*
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

una palabra de código a la vez. De manera específica, éste conforma lo que es equivalente a un *árbol de decisión*, el cual es una representación gráfica de las palabras del código en el código fuente particular. Por ejemplo, la figura 9.4 describe el árbol de decisión correspondiente al código II en la tabla 9.2. El árbol tiene un *estado inicial* y cuatro *estados terminales* correspondientes a los símbolos fuente s_0, s_1, s_2 y s_3. El decodificador siempre empieza en el estado inicial. El primer bit recibido mueve al decodificador al estado terminal s_0 si éste es 0, u otro a un segundo punto de decisión si éste es 1. En el último caso, el segundo bit mueve al decodificador un paso adicional hacia abajo del árbol, ya sea al estado terminal s_1 si éste es 0 u otro al tercer punto de decisión si éste es 1, etc. Una vez que cada estado terminal emite su símbolo, el decodificador se restablece a su estado inicial. Advierta también que cada bit en la secuencia codificada recibida se examina sólo una vez. Por ejemplo, la secuencia codificada 1011111000 se decodifica de inmediato como la secuencia fuente $s_1s_3s_2s_0s_0$. Se invita al lector a que realice esta decodificación.

Un código de prefijo tiene la importante propiedad de que *siempre* es decodificable en forma única. Sin embargo, lo inverso no es necesariamente cierto. Por ejemplo, el código III en la tabla 9.2 no satisface la condición de prefijo, aunque sea decodificable en forma única, pues el bit 0 indica el principio de cada palabra de código en el código.

Además, si un código de prefijo se ha construido para una fuente discreta sin memoria con alfabeto de fuente $\{s_0, s_1, ..., s_{K-1}\}$ y estadísticas de fuente $\{p_0, p_1, ..., p_{K-1}\}$, y la palabra de código para el símbolo s_k tiene longitud l_k, $k = 0, 1, ..., K - 1$, entonces las longitudes de palabra del código satisfacen siempre una cierta desigualdad conocida como la *desigualdad de Kraft-McMillan*,[5] indicada mediante

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1 \tag{9.22}$$

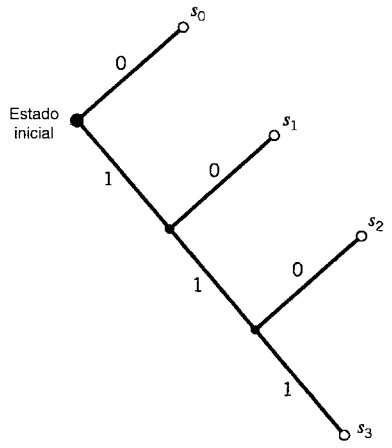


FIGURA 9.4 Árbol de decisiones para el código II de la tabla 9.2.

donde el factor 2 se refiere a la base (número de símbolo) en el alfabeto binario. Sin embargo, resulta importante advertir que la desigualdad de Kraft-McMillan *no* nos indica que un código fuente sea un código de prefijo. Más bien, es meramente una condición sobre las longitudes de palabra de código del código y no sobre las propias palabras de código. Por ejemplo, refiriéndonos a los tres códigos listados en la tabla 9.2, advertimos lo siguiente:

- El código I viola la desigualdad de Kraft-McMillan; por tanto, éste no puede ser un código de prefijo.
- Los códigos II y III satisfacen la desigualdad de Kraft-McMillan; pero sólo el código II es de prefijo.

Los códigos de prefijo se distinguen de otros códigos decodificables en forma única porque siempre es reconocible el final de la palabra de código. Por tanto, la decodificación de un prefijo puede lograrse tan pronto como la secuencia binaria que representa a un símbolo de fuente se recibe completamente. Por esta razón, los códigos de prefijo se conocen también como *códigos instantáneos.*

Dada una fuente discreta sin memoria de entropía $H(\mathscr{S})$, es posible construir un código de prefijo con una longitud promedio de palabra de código $\overline{L}$, el cual está delimitado de la manera siguiente:

$$H(\mathscr{S}) \leq \overline{L} < H(\mathscr{S}) + 1 \tag{9.23}$$

La cota del primer miembro de la ecuación (9.23) se satisface con la igualdad bajo la condición de que el símbolo s_k sea emitido por la fuente con probabilidad

$$p_k = 2^{-l_k} \tag{9.24}$$

donde l_k es la longitud de la palabra de código asignada al símbolo fuente s_k. Tenemos entonces

$$\sum_{k=0}^{K-1} 2^{-l_k} = \sum_{k=0}^{K-1} p_k = 1$$

Bajo esta condición, la desigualdad de Kraft-McMillan de la ecuación (9.22) nos indica que podemos construir un código de prefijo, tal que la longitud de la palabra de código asignada al símbolo de fuente s_k sea $-\log_2 p_k$. Para dicho código, la longitud promedio de palabra es

$$\overline{L} = \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}} \tag{9.25}$$

y la entropía correspondiente de la fuente corresponde a

$$\begin{aligned} H(\mathscr{S}) &= \sum_{k=0}^{K-1} \left(\frac{1}{2^{l_k}} \right) \log_2 (2^{l_k}) \\ &= \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}} \end{aligned} \tag{9.26}$$

Por tanto, en este caso especial (más bien engañoso), encontramos a partir de las ecuaciones (9.25) y (9.26) que el código de prefijo se *acopla* a la fuente en tanto que $\overline{L} = H(\mathscr{S})$.

¿Pero cómo acoplamos el código de prefijo a una fuente discreta sin memoria arbitraria? La respuesta a este problema se ubica en el uso de un *código extendido.* Dejemos que $\overline{L}_n$ denote la longitud promedio de palabra de código del código de prefijo extendido. Para un código decodificable en forma única, $\overline{L}_n$ es la más pequeña posible. De acuerdo con la ecuación (9.23), deducimos que

$$H(\mathscr{S}^n) \leq \overline{L}_n < H(\mathscr{S}^n) + 1 \tag{9.27}$$

Al sustituir la ecuación (9.17) para una fuente extendida en la ecuación (9.27), obtenemos

$$nH(\mathscr{S}) \leq \overline{L}_n < nH(\mathscr{S}) + 1$$

o, de manera equivalente,

$$H(\mathscr{S}) \leq \frac{\overline{L}_n}{n} < H(\mathscr{S}) + \frac{1}{n} \tag{9.28}$$

En el límite, cuando n tiende a infinito, las cotas inferior y superior de la ecuación (9.28) convergen, según indica

$$\lim_{n \to \infty} \frac{1}{n} \overline{L}_n = H(\mathscr{S}) \tag{9.29}$$

Por tanto, es posible establecer que al hacer el orden n de un codificador de fuente de prefijo extendido lo suficientemente grande, podemos lograr que el código represente fielmente la fuente discreta sin memoria $\mathscr{S}$ con la precisión que se desee. En otras palabras, la longitud promedio de la palabra de código de un código de prefijo extendido puede hacerse tan pequeña como la entropía de la fuente, siempre que el código extendido tenga un orden lo bastante elevado, de acuerdo con el teorema de la codificación de la fuente. Sin embargo, el precio que debe pagarse para reducir la longitud promedio de la palabra de código es una mayor complejidad en la decodificación, lo cual es una consecuencia del elevado orden del código de prefijo extendido.

■ CODIFICACIÓN HUFFMAN

A continuación describimos una clase importante de códigos de prefijo conocidos como códigos Huffman. La idea básica detrás de la *codificación Huffman*[6] es asignar a cada símbolo del alfabeto una secuencia de bits aproximadamente igual en longitud a la cantidad de información que transporta el símbolo correspondiente. El resultado final es un código fuente cuya longitud promedio de palabra se acerca al límite fundamental fijado por la entropía de una fuente discreta sin memoria, a saber, $H(\mathscr{S})$. La esencia del *algoritmo* utilizado para sintetizar el código Huffman es sustituir el conjunto preestablecido de estadísticas de una fuente discreta sin memoria por uno más simple. Este proceso de *reducción* se continúa paso por paso hasta que nos quedamos al final con un conjunto de únicamente dos estadísticas (símbolos) de fuente, para las cuales (0, 1) es un código óptimo. Partiendo de este código trivial, trabajamos entonces hacia atrás y construimos de este modo el código Huffman para la fuente dada.

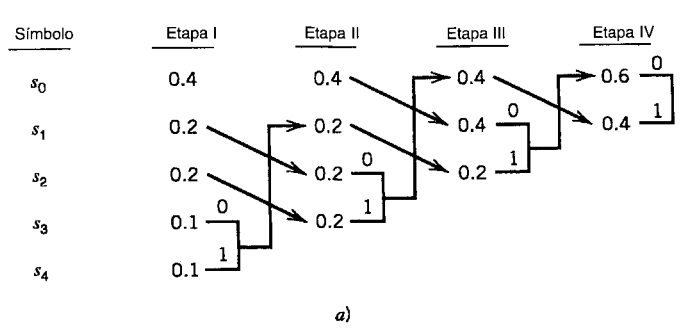
De modo específico, el *algoritmo de codificación* Huffman procede de la manera siguiente:

1. Los símbolos de la fuente se listan en orden de probabilidad decreciente. A los dos símbolos de fuente de probabilidad más baja se les asigna un 0 y un 1. Esta parte del proceso se conoce como etapa de *división*.
2. Estos dos símbolos de fuente se consideran como si se combinaran en un nuevo símbolo de fuente con probabilidad igual a la suma de las dos probabilidades originales. (La lista de símbolos de fuente, y por tanto las estadísticas de la fuente, reducen por ello su tamaño en una unidad.) La probabilidad del nuevo símbolo se pone en la lista de acuerdo con su valor.
3. El procedimiento se repite hasta que nos quedamos con la lista final de las dos únicas estadísticas (símbolos) a las cuales se asignaron un 0 y un 1.

El código para cada símbolo de fuente (original) se determina procediendo hacia atrás y siguiendo la frecuencia de ceros y unos asignada a ese símbolo, así como a sus sucesores.

▶ EJEMPLO 9.3 Árbol de Huffman

Los cinco símbolos del alfabeto de una fuente discreta sin memoria y sus probabilidades se muestran en las dos columnas más a la izquierda de la figura 9.5*a*. Siguiendo el algoritmo Huffman, llegamos al fin del cálculo en cuatro pasos, lo que da origen al *árbol de Huffman* que se presenta en la figura 9.5*a*. Las palabras del código Huffman para la fuente se tabulan en la figura 9.5*b*.



Símbolo	Probabilidad	Palabra de código
s_0	0.4	00
s_1	0.2	10
s_2	0.2	11
s_3	0.1	010
s_4	0.1	011

b)

FIGURA 9.5 *a*) Ejemplo del algoritmo de codificación Huffman. *b*) Código fuente.

La longitud promedio de palabra de código es entonces

$$\begin{aligned}\bar{L} &= 0.4(2)+0.2(2)+0.2(2)+0.1(3)+0.1(3)\\ &= 2.2\end{aligned}$$

La entropía de esta fuente discreta sin memoria especificada se calcula de la manera siguiente [vea la ecuación (9.9)]:

$$\begin{aligned}H(\mathscr{S}) &= 0.4\log_2\left(\frac{1}{0.4}\right)+0.2\log_2\left(\frac{1}{0.2}\right)+0.2\log_2\left(\frac{1}{0.2}\right)\\ &\quad +0.1\log_2\left(\frac{1}{0.1}\right)+0.1\log_2\left(\frac{1}{0.1}\right)\\ &= 0.52877+0.46439+0.46439+0.33219+0.33219\\ &= 2.12193 \text{ bits}\end{aligned}$$

En el ejemplo presente, hacemos las dos observaciones:

1. La longitud promedio de palabra de código $\bar{L}$ supera a la entropía $H(\mathscr{S})$ por únicamente 3.67%.
2. La longitud promedio de palabra de código $\bar{L}$ satisface en realidad la ecuación (9.23). ◀

Conviene señalar que el proceso de codificación Huffman (es decir, el árbol de Huffman) no es único. En particular, es posible citar dos variaciones en el proceso que son responsables de la no unicidad del código Huffman. En primer lugar, en cada etapa de división en la construcción del código Huffman hay arbitrariedad en la manera en que se asignan un 0 y un 1 a los últimos dos símbolos de fuente. Sin embargo, cualquiera que sea la forma en la que se realizan las asignaciones, las diferencias que se producen son triviales. En segundo lugar, surge ambigüedad cuando la probabilidad de un símbolo *combinado* (obtenido al sumar las últimas dos probabilidades pertinentes a un paso particular) se encuentra igual a otra probabilidad en la lista. Podemos proseguir ubicando la probabilidad del nuevo símbolo tan *alta* como sea posible, al igual que en el ejemplo 9.3. De modo alternativo, podemos situarla tan *baja* como sea posible. (Se supone que cualquiera que sea el lugar que se considere, alto o bajo, éste se adhiere consistentemente a lo largo del proceso de codificación.) Pero esta vez se produce una diferencia notable en el sentido de que las palabras del código de fuente resultante pueden tener longitudes diferentes. A pesar de eso, la longitud promedio de la palabra de código permanece igual.

Como una medida de la variabilidad en las longitudes de palabra de código de un código fuente, definimos la *varianza* de la longitud promedio de palabra de código $\bar{L}$ sobre el conjunto de símbolos fuente como

$$\sigma^2 = \sum_{k=0}^{K-1} p_k(l_k - \bar{L})^2 \tag{9.30}$$

donde $p_0, p_1, \ldots, p_{K-1}$ son las estadísticas de fuente y l_k es la longitud de la palabra de código asignada al símbolo fuente s_k. Suele observarse que cuando un símbolo encontrado se mueve lo

más alto posible, el código de Huffman resultante tiene una varianza σ^2 bastante más pequeña que cuando se mueve lo más bajo posible. Con base en esto, resulta razonable elegir el primer código de Huffman en lugar del segundo.

■ CODIFICACIÓN LEMPEL-ZIV

Una desventaja del código de Huffman es que requiere conocimiento de un modelo probabilístico de la fuente; infortunadamente, en la práctica, las estadísticas de fuente no siempre se conocen *a priori*. Además, al modelar texto descubrimos que los requerimientos de almacenamiento evitan que el código de Huffman capture las relaciones de orden superior entre palabras y frases, comprometiendo por ello la eficiencia de dicho código. Para superar estas limitaciones prácticas, es posible recurrir al *algoritmo de Lempel-Ziv*,[7] el cual es intrínsecamente *adaptable* y más simple de poner en práctica que la codificación Huffman.

Básicamente, la codificación en el algoritmo Lempel-Ziv se lleva a cabo descomponiendo el flujo de datos fuente en segmentos que son las subsecuencias más cortas no encontradas previamente. Para ilustrar esta simple pero elegante idea, considere el ejemplo de una secuencia binaria de entrada especificada como sigue:

000101110010100101 . . .

Se supone que los símbolos binarios 0 y 1 están realmente almacenados en ese orden en el registro de códigos. En ese caso escribimos

Subsecuencias almacenadas: 0, 1
Datos por analizar: 000101110010100101. . .

El proceso de codificación se inicia a la izquierda. Con los símbolos 0 y 1 ya almacenados, la *subsecuencia más corta* del flujo de datos encontrados en la primera ocasión y no vista antes es 00; por ello escribimos

Subsecuencias almacenadas: 0, 1, 00
Datos por analizar: 0101110010100101 . . .

La segunda subsecuencia más corta no vista antes es 01; por tanto, escribimos

Subsecuencias almacenadas: 0, 1, 00, 01
Datos por analizar: 01110010100101 . . .

La siguiente subsecuencia más corta no encontrada antes es 011; por tanto, escribimos

Subsecuencias almacenadas: 0, 1, 00, 01, 011
Datos por analizar: 10010100101 . . .

Continuamos en la manera descrita aquí hasta analizar por completo el flujo de datos dado. De esa manera, para el ejemplo presente obtenemos el *registro de códigos* de las subsecuencias binarias que se muestran en el segundo renglón de la figura 9.6.

Posiciones numéricas:	1	2	3	4	5	6	7	8	9
Subsecuencias:	0	1	00	01	011	10	010	100	101
Representaciones numéricas:			11	12	42	21	41	61	62
Bloques codificados binarios:			0010	0011	1001	0100	1000	1100	1101

FIGURA 9.6 Ilustración de los procesos de codificación efectuados por el algoritmo Lempel-Ziv sobre la secuencia binaria 000101110010100101 ...

El primer renglón que se muestra en esta figura indica meramente las posiciones numéricas de las subsecuencias individuales en el registro de código. Después de esto advertimos que la primera subsecuencia del flujo de datos, 00, está conformada por la concatenación consigo misma de la *primera* entrada del registro de código, 0; por tanto está representada por el número 11. La segunda subsecuencia del flujo de datos, 01, consiste en la *primera* entrada del registro de código, 0, concatenada con la *segunda* entrada del registro de código, 1; en consecuencia, se representa mediante el número 12. Las subsecuencias restantes se tratan de manera similar. El conjunto complejo de representaciones numéricas para las diversas subsecuencias en el registro de código se muestra en el tercer renglón de la figura 9.6. Como un ejemplo adicional que ilustra la composición de este renglón, notamos que la subsecuencia 010 está compuesta por la concatenación de la subsecuencia 01 en la posición 4 y el símbolo 0 en la posición 1; de ahí la representación numérica 41. El último renglón que se muestra en la figura 9.6 es la representación codificada binaria de las diferentes subsecuencias del flujo de datos.

El último símbolo de cada subsecuencia en el registro de códigos (es decir, el segundo renglón de la figura 9.6) es un *símbolo de innovación*, el cual se nombra de ese modo en reconocimiento de que su pertenencia a una subsecuencia particular lo distingue de todas las subsecuencias previas almacenadas en el registro de código. De modo correspondiente, el último bit de cada bloque uniforme de bits en la representación codificada binaria del flujo de datos (es decir, el cuarto renglón en la figura 9.6) representa al símbolo de innovación para la subsecuencia particular que se está considerando. Los bits restantes proporcionan la representación binaria equivalente del "indicador" correspondiente a la *subsecuencia de raíz* que asocia a aquella en cuestión salvo por el símbolo de innovación.

El decodificador es igual de simple que el codificador. Específicamente, utiliza el indicador para identificar la subsecuencia raíz y luego agrega el símbolo de innovación. Considere, por ejemplo, el registro codificado binario 1101 en la posición 9. El último bit, 1, es el símbolo de innovación. Los 110 bits restantes indican la subsecuencia de raíz 10 en la posición 6. Por tanto, el bloque 1101 se decodifica en 101, lo cual es correcto.

A partir del ejemplo descrito aquí, advertimos que, en contraste con la codificación Huffman, el algoritmo Lempel-Ziv utiliza códigos de longitud fija para representar un número variable de símbolos fuente; este rasgo hace que el código Lempel-Ziv sea adecuado para la transmisión síncrona. En la práctica se usan bloques fijos de 12 bits de longitud, lo que implica un registro de código de 4096 entradas.

Durante mucho tiempo, la codificación Huffman no tuvo rival como el algoritmo favorito para la compactación de datos. Sin embargo, el algoritmo Lempel-Ziv ha ocupado el puesto casi por completo del algoritmo Huffman y es ahora el estándar para la compresión de archivos. Cuando se aplica a un texto ordinario en inglés, el algoritmo Lempel-Ziv alcanza una compactación de aproximadamente 55%. Esto se compara con una compactación de casi 43% lograda con la codificación Huffman. La razón de este comportamiento es que, como se mencionó antes, la codificación Huffman no aprovecha las redundancias intercaracteres del lenguaje. Por otra parte, el algoritmo Lempel-Ziv es capaz de realizar la mejor compactación posible del texto (dentro de ciertos límites) trabajando efectivamente a niveles superiores.

9.5 *Canales discretos sin memoria*

Hasta ahora en el capítulo nos hemos preocupado de las fuentes sin memoria discretas responsables de la generación de información. A continuación consideraremos el aspecto de la transmisión de información, destacando en particular la confiabilidad. Iniciamos la exposición considerando un canal discreto sin memoria, la contraparte de una fuente discreta sin memoria.

Un *canal discreto sin memoria* es un modelo estadístico con una entrada X y una salida Y que es una versión *ruidosa* de X. Tanto X como Y son variables aleatorias. Cada unidad de tiempo, el

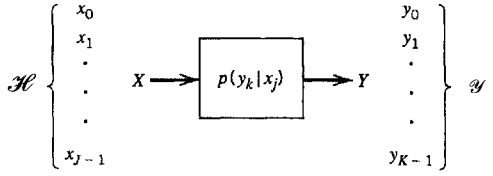


FIGURA 9.7 Canal discreto sin memoria.

canal acepta un símbolo de entrada X elegido de un alfabeto $\mathscr{H}$ y, en respuesta emite un símbolo Y de un alfabeto $\mathscr{Y}$. Se dice que el canal será "discreto" cuando ambos alfabetos $\mathscr{H}$ y $\mathscr{Y}$ tienen tamaños *finitos*. Se dice que será "sin memoria" cuando el símbolo de salida presente dependa *exclusivamente* del símbolo de entrada presente y *no* de cualquier símbolo previo.

La figura 9.7 presenta una descripción de un canal discreto sin memoria. El canal se describe en términos de un *alfabeto de entrada*

$$\mathscr{H} = \{x_0, x_1, \ldots, x_{J-1}\}, \tag{9.31}$$

un *alfabeto de salida,*

$$\mathscr{Y} = \{y_0, y_1, \ldots, y_{K-1}\}, \tag{9.32}$$

y un conjunto de *probabilidades de transición*

$$p(y_k|x_j) = P(Y = y_k|X = x_j) \quad \text{para toda } j \text{ y } k \tag{9.33}$$

Naturalmente, tenemos

$$0 \le p(y_k|x_j) \le 1 \quad \text{para toda } j \text{ y } k \tag{9.34}$$

Además, el alfabeto de entrada $\mathscr{H}$ y el alfabeto de salida $\mathscr{Y}$ no necesitan tener el mismo tamaño. Por ejemplo, en la codificación de canal, el tamaño K del alfabeto de salida $\mathscr{Y}$ puede ser mayor que el tamaño J del alfabeto de entrada $\mathscr{H}$; por tanto, $K \ge J$. Por otra parte, es posible tener una situación en la que el canal emita el mismo símbolo cuando se envíe uno de dos símbolos de entrada, en cuyo caso tendremos $K \le J$.

Una manera conveniente de describir un canal discreto sin memoria consiste en arreglar las diversas probabilidades de transición del canal en la forma de una matriz:

$$\mathbf{P} = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) & \cdots & p(y_{K-1}|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \cdots & p(y_{K-1}|x_1) \\ \vdots & \vdots & & \vdots \\ p(y_0|x_{J-1}) & p(y_1|x_{J-1}) & \cdots & p(y_{K-1}|x_{J-1}) \end{bmatrix} \tag{9.35}$$

La matriz $\mathbf{P}$ de J por K recibe el nombre de *matriz de canal* o *matriz de transición.* Advierta que cada *renglón* de la matriz de canal $\mathbf{P}$ corresponde a una *entrada de canal fija,* en tanto que cada columna de la matriz corresponde a una *salida de canal fija.* Advierta también que una propiedad fundamental del canal de matriz $\mathbf{P}$, según se define aquí, es que la suma de los elementos a lo largo del renglón de la matriz siempre es igual a la unidad; esto es,

$$\sum_{k=0}^{K-1} p(y_k|x_j) = 1 \qquad \text{para toda } j \tag{9.36}$$

Suponga ahora que las entradas a un canal sin memoria discreta se eligen de acuerdo con la *distribución de probabilidad* $\{p(x_j), j = 0, 1, ..., J-1\}$. En otras palabras, el evento de que la entrada del canal $X = x_j$ ocurra con probabilidad

$$p(x_j) = P(X = x_j) \quad \text{para } j = 0, 1, \ldots, J-1 \tag{9.37}$$

Habiendo especificado la variable aleatoria X que denota la entrada del canal, es posible especificar ahora la segunda variable aleatoria Y que denota la salida del canal. La *distribución de probabilidad conjunta* de las variables aleatorias X y Y está dada por

$$\begin{aligned} p(x_j, y_k) &= P(X = x_j, Y = y_k) \\ &= P(Y = y_k | X = x_j) P(X = x_j) \\ &= p(y_k | x_j) p(x_j) \end{aligned} \tag{9.38}$$

La *distribución de probabilidad marginal* de la variable aleatoria de salida Y se obtiene promediando la dependencia de $p(x_j, y_k)$ con respecto a x_j, como indica

$$\begin{aligned} p(y_k) &= P(Y = y_k) \\ &= \sum_{j=0}^{J-1} P(Y = y_k | X = x_j) P(X = x_j) \\ &= \sum_{j=0}^{J-1} p(y_k | x_j) p(x_j) \qquad \text{para } k = 0, 1, \ldots, K-1 \end{aligned} \tag{9.39}$$

Las probabilidades $p(x_j)$ para $j = 0, 1, ..., J-1$, se conocen como las *probabilidades a priori* de los distintos símbolos de entrada. La ecuación (9.39) establece que si se nos dan las probabilidades *a priori* de entrada $p(x_j)$ y la matriz de canal [es decir, la matriz de las probabilidades de transición $p(y_k | x_j)$], es posible calcular entonces las probabilidades de los diversos símbolos de salida, las $p(y_k)$.

▶ EJEMPLO 9.4 Canal simétrico binario

El *canal simétrico binario* es de gran interés teórico e importancia práctica. Constituye un caso especial de canal discreto sin memoria con $J = K = 2$. El canal tiene dos símbolos de entrada ($x_0 = 0$, $x_1 = 1$) y dos símbolos de salida ($y_0 = 0$, $y_1 = 1$); es simétrico porque la probabilidad de recibir un 1 si se envió un 0 es la misma que la probabilidad de recibir un 0 si se envió un 1. Esta probabilidad condicional de error se denota por medio de p. El *diagrama de probabilidad de transición* de un canal simétrico binario es como se muestra en la figura 9.8. ◀

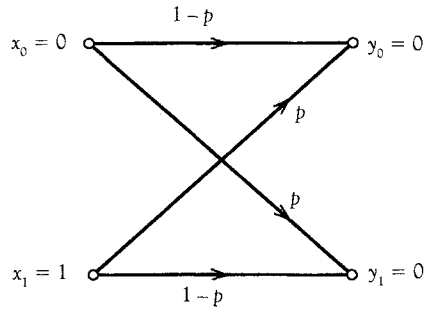


FIGURA 9.8 Diagrama de probabilidad de transición de un canal simétrico binario.

Es de interés relacionar el diagrama de probabilidad de transición de la figura 9.8 con las probabilidades condicionales de error p_{10} y p_{01} que se determinaron para el receptor PCM en la sección 3.3. Para el caso en el que los símbolos 0 y 1 son igualmente probables, mostramos que los valores optimizados de estas dos probabilidades de error son iguales. En realidad, al recordar las definiciones siguientes (utilizando la terminología de la figura 9.8):

$$p_{10} = p(y = 1|x = 0)$$

y

$$p_{01} = P(y = 0|x = 1)$$

advertimos de inmediato que para el receptor PCM de la figura 3.4:

$$p_{10} = p_{01} = p$$

9.6 *Información mutua*

Debido a que consideramos la salida del canal Y (elegida del alfabeto $\mathscr{Y}$) como una versión ruidosa de la entrada X del canal (elegida del alfabeto $\mathscr{X}$), y que la entropía $H(\mathscr{X})$ es una medida de la incertidumbre anterior en torno a X, ¿cómo podemos medir la incertidumbre alrededor de X después de observar Y? Para responder esta pregunta, extendemos las ideas formuladas en la sección 9.2 definiendo la *entropía condicional* de X elegida del alfabeto $\mathscr{X}$, dado que $Y = y_k$. De manera específica, escribimos

$$H(\mathscr{X} \mid Y = y_k) = \sum_{j=0}^{J-1} p(x_j \mid y_k) \log_2\left[\frac{1}{p(x_j|y_k)}\right] \tag{9.40}$$

Esta misma cantidad es una variable aleatoria que toma los valores $H(\mathscr{X}|Y = y_0), \ldots,$ $H(\mathscr{X}|Y = y_{K-1})$ con probabilidades $p(y_0), \ldots, p(y_{K-1})$, respectivamente. La media de la entropía $H(\mathscr{X}|Y = y_k)$ sobre el alfabeto de salida Y está dada entonces por

$$\begin{aligned} H(\mathscr{X}|\mathscr{Y}) &= \sum_{k=0}^{K-1} H(\mathscr{X}|Y = y_k)p(y_k) \\ &= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j|y_k)p(y_k) \log_2\left[\frac{1}{p(x_j|y_k)}\right] \\ &= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k) \log_2\left[\frac{1}{p(x_j|y_k)}\right] \end{aligned} \tag{9.41}$$

donde, en la última línea, hemos recurrido a la relación

$$p(x_j, y_k) = p(x_j|y_k)p(y_k) \tag{9.42}$$

La cantidad $H(\mathscr{X}|\mathscr{Y})$ recibe el nombre de *entropía condicional* y representa *la cantidad de incertidumbre que queda acerca de la entrada del canal después de que se ha observado la salida del canal.*

Puesto que la entropía de $H(\mathscr{X})$ representa nuestra incertidumbre en torno a la entrada del canal *antes* de observar la salida del mismo, y la entropía $H(\mathscr{X}|\mathscr{Y})$ representa nuestra incertidumbre con respecto a la entrada del canal *después* de observar la salida de éste, se concluye que la diferencia $H(\mathscr{X}) - H(\mathscr{X}|\mathscr{Y})$ debe representar nuestra incertidumbre en torno a la entrada del

canal que se *resuelve* al observar la salida del mismo. Esta importante cantidad se denomina la *información mutua* del canal. Denotando la información mutua como $I(\mathscr{H}; \mathscr{Y})$, podemos escribir

$$I(\mathscr{H};\mathscr{Y}) = H(\mathscr{H}) - H(\mathscr{H}|\mathscr{Y}) \tag{9.43}$$

De manera similar, expresamos

$$I(\mathscr{Y};\mathscr{H}) = H(\mathscr{Y}) - H(\mathscr{Y}|\mathscr{H}) \tag{9.44}$$

donde $H(\mathscr{Y})$ es la entropía de la salida del canal y $H(\mathscr{Y}|\mathscr{H})$ es la entropía condicional de la salida del canal dada la entrada del mismo.

■ Propiedades de la información mutua

La información mutua $I(\mathscr{H};\mathscr{Y})$ tiene las siguientes propiedades importantes.

Propiedad 1

La información mutua de un canal es simétrica; esto es

$$I(\mathscr{H};\mathscr{Y}) = I(\mathscr{Y};\mathscr{H}) \tag{9.45}$$

donde la información mutua $I(\mathscr{H};\mathscr{Y})$ es una medida de la incertidumbre en torno a la entrada del canal que se resuelve *observando* la salida del mismo, y la información mutua $I(\mathscr{Y}; \mathscr{H})$ constituye una medida de la incertidumbre en torno de la salida del canal que se resuelve *enviando* la entrada del mismo.

Para probar esta propiedad, utilizamos primero la fórmula para la entropía y después las ecuaciones (9.36) y (9.38), en ese orden, para expresar $H(\mathscr{H})$ como

$$\begin{aligned} H(\mathscr{H}) &= \sum_{j=0}^{J-1} p(x_j) \log_2\left[\frac{1}{p(x_j)}\right] \\ &= \sum_{j=0}^{J-1} p(x_j) \log_2\left[\frac{1}{p(x_j)}\right] \sum_{k=0}^{K-1} p(y_k|x_j) \\ &= \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(y_k|x_j)p(x_j) \log_2\left[\frac{1}{p(x_j)}\right] \\ &= \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k) \log_2\left[\frac{1}{p(x_j)}\right] \end{aligned} \tag{9.46}$$

Por tanto, al sustituir las ecuaciones (9.41) y (9.46) en la (9.43) y combinando después los términos, obtenemos

$$I(\mathscr{H};\mathscr{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k) \log_2\left[\frac{p(x_j|y_k)}{p(x_j)}\right] \tag{9.47}$$

De acuerdo con la *regla de Bayes* para probabilidades condicionales, tenemos [vea las ecuaciones (9.38) y (9.42)]

$$\frac{p(x_j|y_k)}{p(x_j)} = \frac{p(y_k|x_j)}{p(y_k)} \tag{9.48}$$

De ese modo, al sustituir la ecuación (9.48) en la (9.47) e intercambiar el orden de la sumatoria, podemos escribir

$$\begin{aligned} I(\mathscr{X};\mathscr{Y}) &= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j,y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right] \\ &= I(\mathscr{Y};\mathscr{X}) \end{aligned} \tag{9.49}$$

que es el resultado deseado.

Propiedad 2

La información mutua es siempre no negativa; es decir,

$$I(\mathscr{X};\mathscr{Y}) \geq 0 \tag{9.50}$$

Para confirmar esta propiedad, advertimos primero de la ecuación (9.42) que

$$p(x_j|y_k) = \frac{p(x_j,y_k)}{p(y_k)} \tag{9.51}$$

Por tanto, al sustituir la ecuación (9.51) en la (9.47), es posible expresar la información mutua del canal como

$$I(\mathscr{X};\mathscr{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k)\log_2\left(\frac{p(x_j,y_k)}{p(x_j)p(y_k)}\right) \tag{9.52}$$

Después de esto, la aplicación directa de la desigualdad fundamental [definida por la ecuación (9.12)] se produce por el resultado deseado

$$I(\mathscr{X};\mathscr{Y}) \geq 0$$

con la igualdad si, y solamente si,

$$p(x_j,y_k) = p(x_j)p(y_k) \qquad \text{para toda } j \text{ y } k \tag{9.53}$$

La propiedad 2 establece que *no podemos perder información, en promedio, observando la salida de un canal.* Además, la información mutua es cero si, y solamente si, los símbolos de entrada y salida del canal son estadísticamente independientes, como en la ecuación (9.53).

Propiedad 3

La información mutua de un canal se relaciona con la entropía conjunta de la entrada y la salida del mismo mediante

$$I(\mathscr{X};\mathscr{Y}) = H(\mathscr{X}) + H(\mathscr{Y}) - H(\mathscr{X},\mathscr{Y}) \tag{9.54}$$

donde la entropía conjunta $H(\mathscr{X},\mathscr{Y})$ *está definida por*

$$H(\mathscr{X},\mathscr{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k)\log_2\left(\frac{1}{p(x_j,y_k)}\right) \tag{9.55}$$

Para demostrar la ecuación (9.54), describimos primero la definición de la entropía conjunta $H(\mathscr{X},\mathscr{Y})$ como

$$\begin{aligned} H(\mathscr{X},\mathscr{Y}) &= \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k)\log_2\left[\frac{p(x_j)p(y_k)}{p(x_j,y_k)}\right] \\ &\quad + \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k)\log_2\left[\frac{1}{p(x_j)p(y_k)}\right] \end{aligned} \tag{9.56}$$

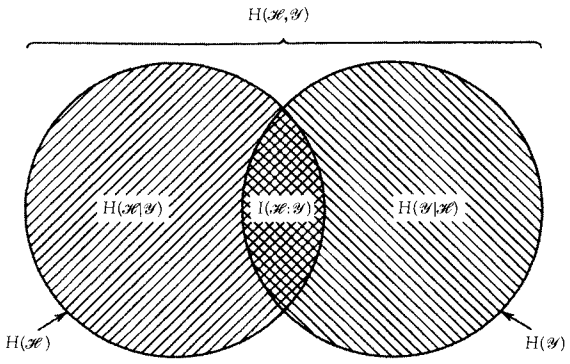


FIGURA 9.9 Ilustración de las relaciones entre diversas entropías de canal.

El primer término de doble sumatoria en el segundo miembro de la ecuación (9.56) se reconoce como el negativo de la información mutua del canal, $I(\mathscr{X},\mathscr{Y})$, dada anteriormente en la ecuación (9.52). En cuanto al segundo término de sumatorias, lo manipulamos de la manera que sigue:

$$\begin{aligned}\sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j,y_k)\log_2\left[\frac{1}{p(x_j)p(y_k)}\right] &= \sum_{j=0}^{J-1}\log_2\left[\frac{1}{p(x_j)}\right]\sum_{k=0}^{K-1} p(x_j,y_k)\\ &\quad+\sum_{k=0}^{K-1}\log_2\left[\frac{1}{p(y_k)}\right]\sum_{j=0}^{J-1} p(x_j,y_k)\\ &= \sum_{j=0}^{J-1} p(x_j)\log_2\left[\frac{1}{p(x_j)}\right]\\ &\quad+\sum_{k=0}^{K-1} p(y_k)\log_2\left[\frac{1}{p(y_k)}\right]\\ &= H(\mathscr{X})+H(\mathscr{Y})\end{aligned} \tag{9.57}$$

Por tanto, utilizando las ecuaciones (9.52) y (9.57) en la ecuación (9.56), obtenemos el resultado

$$H(\mathscr{X},\mathscr{Y}) = -I(\mathscr{X};\mathscr{Y}) + H(\mathscr{X}) + H(\mathscr{Y}) \tag{9.58}$$

Al reordenar términos en esta ecuación, llegamos al resultado que indica la ecuación (9.54), confirmando de esa manera la propiedad 3.

Terminamos nuestra exposición de la información mutua de un canal dando una interpretación diagramática de las ecuaciones (9.43), (9.44) y (9.54). La interpretación se ofrece en la figura 9.9. La entropía de la entrada X del canal se representa mediante el círculo de la izquierda y la correspondiente a la salida Y del canal, por medio del círculo de la derecha. La información mutua del canal se representa por medio del traslape entre estos dos círculos.

9.7 *Capacidad del canal*

Considere un canal discreto sin memoria con alfabeto de entrada $\mathscr{X}$, alfabeto de salida $\mathscr{Y}$ y probabilidades de transición $p(y_k|x_j)$, donde $j = 0, 1, ..., J-1$ y $k = 0, 1, ..., K-1$. La información mutua del canal se define mediante la primera línea de la ecuación (9.49), la cual se reproduce aquí por conveniencia:

$$I(\mathscr{X};\mathscr{Y}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j,y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right]$$

Advertimos aquí que [vea la ecuación (9.38)]

$$p(x_j, y_k) = p(y_k|x_j)p(x_j)$$

Además, de acuerdo con la ecuación (9.39), tenemos

$$p(y_k) = \sum_{j=0}^{J-1} p(y_k|x_j)p(x_j)$$

Con base en estas tres ecuaciones observamos que es necesario conocer la distribución de probabilidad de entrada $\{p(x_j) | j = 0, 1, ..., J - 1\}$, de manera que tengamos la posibilidad de calcular la información mutua $I(\mathscr{X};\mathscr{Y})$. La información mutua de un canal depende entonces no sólo de este mismo, sino también de la manera en la cual se usa.

La distribución de probabilidad de entrada $\{p(x_j)\}$ es evidentemente independiente del canal. Podemos entonces maximizar la información mutua $I(\mathscr{X};\mathscr{Y})$ del canal con respecto a $\{p(x_j)\}$. Por tanto, *definimos la capacidad del canal de un canal discreto sin memoria como la información mutua máxima* $I(\mathscr{X};\mathscr{Y})$ *en cualquier uso simple del canal (es decir, intervalo de transmisión de señales), donde la maximización es sobre todas las distribuciones de probabilidad de entrada posibles* $\{p(x_j)\}$ *en* $\mathscr{X}$. La capacidad del canal se denota comúnmente por medio de C. De ese modo escribimos

$$C = \max_{\{p(x_j)\}} I(\mathscr{X};\mathscr{Y}) \tag{9.59}$$

La capacidad del canal C se mide en *bits por uso del canal* o *bits por transmisión.*

Advierta que la capacidad del canal C es una función exclusiva de las probabilidades de transición $p(y_k|x_j)$, las cuales definen al canal. El cálculo de C implica la maximización de la información mutua $I(\mathscr{X};\mathscr{X})$ sobre J variable [es decir, las probabilidades de entrada $p(x_0)$, ..., $p(x_{J-1})$] sujeta a dos restricciones:

$$p(x_j) \geq 0 \text{ para toda } j$$

y

$$\sum_{j=0}^{J-1} p(x_j) = 1$$

En general, el problema variacional de determinar la capacidad del canal C es una tarea desafiante.

▶ EJEMPLO 9.5 Canal simétrico binario (reexposición)

Considere otra vez el *canal simétrico binario,* el cual se describe por medio del *diagrama de probabilidad de transición* de la figura 9.8. Este diagrama se define en forma única a partir de la probabilidad condicional de error p.

La entropía $H(X)$ se hace máxima cuando la probabilidad de entrada del canal $p(x_0) = p(x_1) = 1/2$, donde x_0 y x_1 son cada una 0 o 1. La información mutua $I(\mathscr{X};\mathscr{Y})$ se hace máxima de modo similar, por lo que es posible escribir

$$C = I(\mathscr{X};\mathscr{Y})\Big|_{p(x_0)=p(x_1)=1/2}$$

De acuerdo con la figura 9.8 tenemos

$$p(y_0|x_1) = p(y_1|x_0) = p$$

y

$$p(y_0|x_0) = p(y_1|x_1) = 1 - p$$

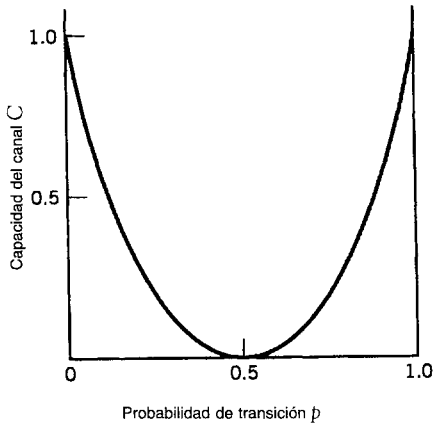


FIGURA 9.10 Variación de la capacidad del canal de un canal simétrico binario con probabilidad de transición p.

Por tanto, sustituyendo estas probabilidades de transición del canal en la ecuación (9.49) con $J = K = 2$, e igualando después la probabilidad de entrada $p(x_0) = p(x_1)$ de acuerdo con la ecuación (9.59), encontramos que la capacidad del canal simétrico binario es

$$C = 1 + p \log_2 p + (1-p) \log_2 (1-p) \tag{9.60}$$

Al utilizar la definición de la función de entropía en la ecuación (9.16) es posible reducir la ecuación (9.60) a

$$C = 1 - H(p)$$

La capacidad del canal C varía con la probabilidad de error (probabilidad de transición) p en una manera convexa, como se muestra en la figura 9.10, la cual es simétrica en torno a $p = 1/2$. Comparando la curva de esta figura con la de la figura 9.2, podemos hacer las siguientes observaciones:

1. Cuando el canal *no tiene ruido*, lo que nos permite dejar $p = 0$, la capacidad C del canal alcanza su valor máximo de un bit por uso del canal, lo cual es exactamente la información en cada entrada del canal. A este valor de p, la función de entropía $H(p)$ llega a su valor mínimo de cero.
2. Cuando la probabilidad condicional de error $p = 1/2$ debido al ruido, la capacidad C del canal alcanza su valor mínimo de cero, en tanto que la función de entropía $H(p)$ llega a su valor máximo de la unidad; en un caso de este tipo se dice que el canal será *inútil*. ◀

9.8 *Teorema de codificación del canal*

La inevitable presencia de *ruido* en un canal provoca discrepancias (errores) entre las secuencias de datos de salida y entrada de un sistema de comunicación digital. En un canal relativamente ruidoso (por ejemplo un canal de comunicación inalámbrico) la probabilidad de error puede alcanzar un valor tan alto como 10^{-1}, lo cual quiere decir que (en promedio) sólo 9 de 10 bits transmitidos se reciben correctamente. En muchas aplicaciones, este *nivel de confiabilidad* resulta inaceptable. En realidad, una probabilidad de error igual a 10^{-6} o incluso menor es muchas veces un requisito necesario. Para conseguir tal nivel de alto desempeño, recurrimos al uso de la codificación del canal.

El objetivo de diseño de la codificación del canal es aumentar la resistencia de un sistema de comunicación digital al ruido del canal. Específicamente, la *codificación del canal* consiste en *hacer corresponder o asignar* la secuencia de datos de entrada con una secuencia de entrada del canal, y en la *correspondencia inversa* de la secuencia de salida del canal con una secuencia de datos de salida, de tal forma que el efecto completo del ruido del canal en el sistema se haga mínimo. La primera operación de correspondencia se realiza en el transmisor por medio de un *codificador de*

canal, donde la operación de correspondencia inversa se lleva a cabo en el receptor a partir de un *decodificador de canal,* como se muestra en el diagrama de bloques de la figura 9.11; para ejemplificar la exposición, no hemos incluido la codificación de fuente (antes de la codificación del canal) y la decodificación de fuente (después de la decodificación del canal) en la figura 9.11.

El codificador de canal y el decodificador de canal en la figura 9.11 se encuentran bajo el control del diseñador y deben diseñarse para optimizar la confiabilidad total del sistema de comunicación. El enfoque que se sigue es introducir *redundancia* en el codificador del canal, de manera que se reconstruya la secuencia fuente original lo más exactamente posible. De esa manera, en un sentido bastante general, es posible considerar la codificación de canal como el *dual* de la codificación de fuente, en el sentido de que la primera introduce redundancia controlada para mejorar la confiabilidad, en tanto que la última reduce la redundancia para mejorar la eficiencia.

El tema de la codificación de canal se trata en el capítulo 10. Para los fines de la explicación presente, es suficiente confinar nuestra atención a los *códigos de bloque.* En esta clase de códigos, la secuencia del mensaje se subdivide en bloques secuenciales cada uno de k bits de largo, y cada bloque de k bits se hace corresponder en un bloque de n bits, donde $n > k$. El número de bits redundantes agregados por el codificador a cada bloque transmitido es igual a $n - k$ bits. La proporción k/n recibe el nombre de *tasa de código.* Utilizando r para denotar la tasa de código, es posible escribir entonces

$$r = \frac{k}{n}$$

donde, desde luego, r es menor que la unidad. Para una k preestablecida, la tasa de código r (y, por tanto, la eficiencia de codificación del sistema) tiende a cero cuando la longitud de bloque n tiende a infinito.

La reconstrucción exacta de la secuencia fuente original en el destino requiere que la *probabilidad promedio del error de símbolo* sea arbitrariamente baja. Esto da lugar a la siguiente pregunta importante: ¿existe un esquema de codificación de canal tal, que la probabilidad de que el bit del mensaje sea erróneo es menor que cualquier número positivo ε (es decir, tan pequeña como deseemos), y que aun así el esquema de codificación de canal sea eficiente en cuanto a que no sea necesario que la tasa de código sea demasiado pequeña? La respuesta a esta pregunta fundamental corresponde a un "sí" definitivo. En realidad, la respuesta a la pregunta la proporciona el segundo teorema de Shannon, en términos de la capacidad del canal C en la forma que se describe más adelante. Hasta este punto, el *tiempo* no ha desempeñado un papel importante en nuestro análisis de la capacidad del canal. Suponga entonces que la fuente discreta sin memoria de la figura 9.11 tiene el alfabeto de fuente $\mathscr{S}$ y entropía $H(\mathscr{S})$ bits por símbolo fuente. Suponemos que la fuente emite símbolos una vez cada T_s segundos. Por tanto, la *tasa de información promedio* de la fuente es $H(\mathscr{S})/T_s$ bits por segundo. El decodificador entrega símbolos decodificados al destino provenientes del alfabeto fuente $\mathscr{S}$ a la misma velocidad de fuente de un símbolo cada T_s segundos. El canal discreto sin memoria tiene una capacidad igual a C bits por uso del canal. Suponemos que el canal es capaz de utilizarse una vez cada T_c segundos. De ese modo, la *capacidad del canal por tiempo unitario es* C/T_c bits por segundo, lo cual representa la tasa máxima de transferencia de información por el canal. Estamos ahora listos para enunciar el segundo teorema de Shannon, conocido como el teorema de codificación del canal.

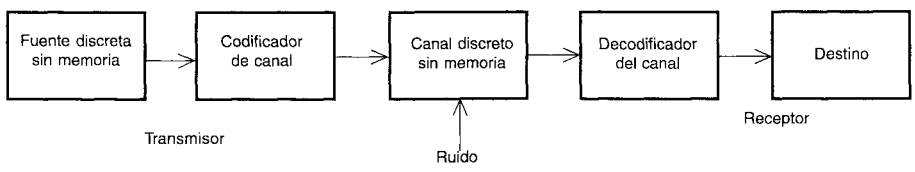


FIGURA 9.11 Diagrama de bloques del sistema de comunicación digital.

De manera específica, el *teorema de codificación del canal*[8] para un canal discreto sin memoria se enuncia en dos partes del modo que sigue.

i) Considere una fuente discreta sin memoria con un alfabeto $\mathscr{S}$ que tiene entropía $H(\mathscr{S})$ y produce símbolos una vez cada T_s segundos. Sea un canal discreto sin memoria con capacidad C y que se usa una vez cada T_c segundos. Entonces, si

$$\frac{H(\mathscr{S})}{T_s} \leq \frac{C}{T_c} \tag{9.61}$$

existe un esquema de codificación para el cual la salida de la fuente puede transmitirse por el canal y reconstruirse con una probabilidad de error arbitrariamente pequeña. El parámetro C/T_c recibe el nombre de *tasa crítica*. Cuando se satisface la ecuación (9.61) con el signo de igualdad, se dice que el sistema transmitirá señales a la tasa crítica.

ii) Inversamente, si

$$\frac{H(\mathscr{S})}{T_s} > \frac{C}{T_c}$$

no es posible transmitir información por el canal y reconstruirla con una probabilidad de error arbitrariamente pequeña.

El teorema de codificación del canal es el resultado simple más importante de la teoría de la información. El teorema especifica la capacidad del canal C como un *límite fundamental* sobre la velocidad a la cual puede ocurrir la transmisión de mensajes confiables sin errores por un canal discreto sin memoria. Sin embargo, es importante advertir lo siguiente:

- El teorema de codificación del canal no nos indica cómo construir un buen código. Más bien, debe considerarse como una *prueba de existencia* porque nos señala que si la condición de la ecuación (9.61) se satisface, entonces existen buenos códigos. (En el capítulo 10 describiremos varios códigos adecuados para los canales discretos sin memoria.)
- El teorema no tiene un resultado preciso para la probabilidad del error de símbolo después de decodificar la salida del canal. En vez de eso, nos indica que la probabilidad del error de símbolo tiende a cero cuando aumenta la longitud del código, siempre y cuando se satisfaga de nuevo la condición de la ecuación (9.61).

Advierta también que las restricciones de potencia y ancho de banda se omitieron en la explicación que se presentó aquí. A pesar de todo, estas dos restricciones del sistema se integran en realidad en la matriz del canal P del canal discreto sin memoria. Esta observación se confirma fácilmente vinculando los resultados del ejemplo 9.5 acerca del canal simétrico binario con el análisis de ruido para el receptor PCM presentado en la sección 5.3.

Aplicación del teorema de la codificación del canal a canales simétricos binarios

Considere una fuente discreta sin memoria que emite símbolos binarios igualmente probables (ceros y unos) cada T_s segundos. Con la entropía de la fuente igual a un bit por símbolo de fuente (vea el ejemplo 9.1), la tasa de información de la fuente es $(1/T_s)$ bits por segundo. La secuencia de fuente se aplica a un codificador de canal con tasa de código r. El codificador del canal produce un símbolo una vez cada T_c segundos. Por tanto, la tasa de transmisión de símbolos codificados es $(1/T_c)$ símbolos por segundo. El codificador del canal ocupa un canal simétrico binario una vez cada T_c segundos. De ese modo, la capacidad del canal por unidad de tiempo es (C/T_s) bits por segundo, donde C se determina mediante la probabilidad de transición del canal

preestablecida p de acuerdo con la ecuación (9.60). Por tanto, el teorema de codificación del canal [parte *i*] implica que si

$$\frac{1}{T_s} \leq \frac{C}{T_c} \tag{9.62}$$

la probabilidad de error puede hacerse arbitrariamente pequeña por medio del uso de un esquema de codificación de canal adecuado. Pero la proporción T_c/T_s es igual a la tasa de código de codificación del canal:

$$r = \frac{T_c}{T_s} \tag{9.63}$$

Por tanto, es posible restablecer la condición de la ecuación (9.62) como

$$r \leq C \tag{9.64}$$

Es decir, para $r \leq C$, existe un código (con tasa de código menor o igual que C) capaz de alcanzar una probabilidad de error arbitrariamente baja.

▶ EJEMPLO 9.6 Código de repetición

En este ejemplo presentamos una interpretación gráfica del teorema de codificación de canal. También traemos a discusión un aspecto sorprendente del teorema al considerar un esquema de codificación simple.

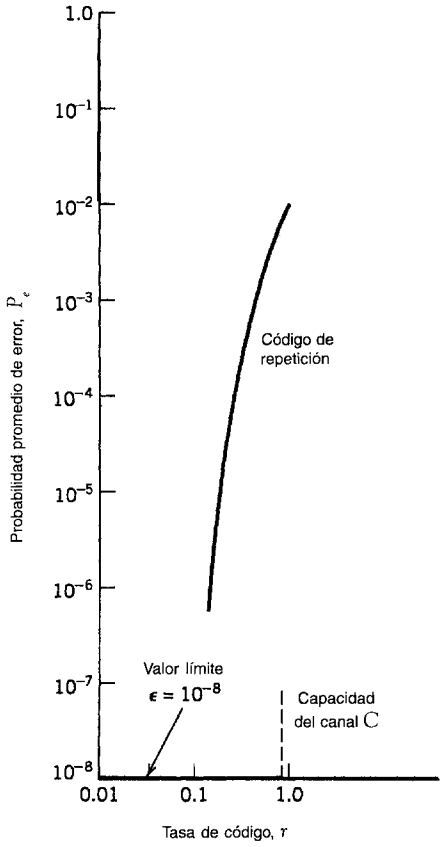


FIGURA 9.12 Ilustración del significado del teorema de la codificación del canal.

Tabla 9.3 ***Probabilidad promedio de error para el código de repetición***

Tasa de código, $r = 1/n$	*Probabilidad promedio de error, P_e*
1	10^{-2}
$\frac{1}{3}$	3×10^{-4}
$\frac{1}{5}$	10^{-6}
$\frac{1}{7}$	4×10^{-7}
$\frac{1}{9}$	10^{-8}
$\frac{1}{11}$	5×10^{-10}

Considere en primer lugar un canal simétrico binario con probabilidad de transición $p = 10^{-2}$. Para este valor de p, encontramos de acuerdo con la ecuación (9.60) que la capacidad del canal $C = 0.9192$. Por tanto, a partir del teorema de codificación del canal, es posible establecer que para cualquier $\varepsilon > 0$ y $r \leq 0.9192$, existe un código de longitud n suficientemente grande y una tasa de código r y un algoritmo de decodificación apropiado, tales que cuando la secuencia de bits codificado se envía por el canal dado, la probabilidad de error de codificación del canal es menor que ε. Este resultado se describe en la figura 9.12 para el valor límite $\varepsilon = 10^{-8}$.

Para poner en perspectiva la importancia de este resultado, considere a continuación un sistema de codificación simple que implica el uso de un *código de repetición*, en el cual cada bit del mensaje se repite varias veces. Considere que cada bit (0 o 1) se repite n veces, donde $n = 2m + 1$ es un entero impar. Por ejemplo, para $n = 3$ transmitimos 0 y 1 como 000 y 111, respectivamente. En forma intuitiva, parecería lógico utilizar una *regla de mayoría* para la decodificación, la cual opera en la forma siguiente: *si en un bloque de n bits recibidos (representando un bit de mensaje), el número de ceros supera al número de unos, el decodificador decide en favor de 0. En otro caso, decide en favor de 1.* Por tanto, ocurre un error cuando $m + 1$ o más bits fuera de los $n = 2m + 1$ bits se reciben incorrectamente. Debido a la naturaleza simétrica supuesta del canal, la *probabilidad promedio de error* P_e es independiente de las probabilidades *a priori* de 0 y 1. Por tanto, encontramos que P_e está dada por (vea el problema 9.24)

$$P_e = \sum_{i=m+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{9.65}$$

donde p es la probabilidad de transición del canal.

La tabla 9.3 proporciona la probabilidad promedio de error P_e para un código de repetición, la cual se calcula utilizando la ecuación (9.65) con respecto a diferentes valores de la tasa de código r. Los valores que se dan aquí suponen el uso de un canal simétrico binario con probabilidad de transición $p = 10^{-2}$. La mejora en la confiabilidad que se exhibe en la tabla 9.3 se consigue a costa de una tasa de código decreciente. Los resultados de esta tabla se presentan también graficados como la curva denominada "código de repetición" en la figura 9.12, la cual ilustra el *intercambio de tasa de código por confiabilidad del mensaje*, lo que constituye una característica de los códigos de repetición.

Este ejemplo subraya el inesperado resultado que nos presenta el teorema de codificación del canal: no es necesario tener una tasa de código r que tienda a cero (como en el caso de los códigos de repetición) a fin de conseguir una operación más y más confiable del enlace de comunicación. El teorema meramente requiere que la tasa de código sea menor que la capacidad C del canal. ◀

9.9 *Entropía diferencial e información mutua para conjuntos continuos*

Las fuentes y canales consideradas hasta ahora en nuestra exposición de los conceptos teóricos de la información han incluido conjuntos de variables aleatorias que son *discretos* en amplitud. En

esta sección extendemos algunos de estos conceptos a variables aleatorias *continuas* y a vectores aleatorios. La motivación en torno a ello es allanar el camino para la descripción de otro límite fundamental en la teoría de la información, mismo que abordaremos en la sección 9.10.

Considere una variable aleatoria continua X con la *función de densidad de probabilidad* $f_X(x)$. Por analogía con la entropía de una variable aleatoria discreta, introducimos la siguiente definición:

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2\left[\frac{1}{f_X(x)}\right] dx \tag{9.66}$$

Nos referimos a $h(X)$ como la *entropía diferencial de* X para distinguirla de la *entropía absoluta* u ordinaria. Hacemos lo anterior en reconocimiento de que aunque $h(X)$ es una cantidad matemáticamente útil que conviene conocer, *no* es en ningún sentido una medida de la aleatoriedad de X. No obstante, justificamos el uso de la ecuación (9.66) del siguiente modo. Empezamos considerando la variable aleatoria X como la forma límite de una variable aleatoria discreta que toma los valores $x_k = k\,\Delta x$, donde $k = 0, \pm 1, \pm 2, \ldots$, y Δx tiende a cero. Por definición, la variable aleatoria continua X toma un valor en el intervalo $[x_k, x_k + \Delta x]$ con probabilidad $f_X(x_k)\,\Delta x$. Por tanto, al dejar que Δx tienda a cero, la entropía ordinaria de la variable aleatoria continua X puede escribirse en el límite como sigue:

$$\begin{aligned} H(X) &= \lim_{\Delta x \to 0} \sum_{k=-\infty}^{\infty} f_X(x_k)\,\Delta x \log_2\left(\frac{1}{f_X(x_k)\,\Delta x}\right) \\ &= \lim_{\Delta x \to 0}\left[\sum_{k=-\infty}^{\infty} f_X(x_k) \log_2\left(\frac{1}{f_X(x_k)}\right)\Delta x - \log_2 \Delta x \sum_{k=-\infty}^{\infty} f_X(x_k)\,\Delta x\right] \\ &= \int_{-\infty}^{\infty} f_X(x) \log_2\left(\frac{1}{f_X(x)}\right) dx - \lim_{\Delta x \to 0} \log_2 \Delta x \int_{-\infty}^{\infty} f_X(x)\,dx \\ &= h(X) - \lim_{\Delta x \to 0} \log_2 \Delta x \end{aligned} \tag{9.67}$$

donde, en la última línea, hemos recurrido a la ecuación (9.66) y al hecho de que el área total bajo la curva de la función de densidad de probabilidad $f_X(x)$ es la unidad. En el límite, cuando Δx tiende a cero, $-\log_2 \Delta x$ tiende a infinito. Esto quiere decir que la entropía de una variable aleatoria continua es infinitamente grande. De manera intuitiva, esperaríamos que esto fuera cierto, debido a que una variable aleatoria continua puede tomar cualquier valor en el intervalo $(-\infty, \infty)$ y la incertidumbre asociada con la variable es del orden de infinito. Evitamos el problema asociado con el término $\log_2 \Delta x$ adoptando $h(X)$ como una entropía diferencial, con el término $-\log_2 \Delta x$ sirviendo como referencia. Además, puesto que la información transmitida por un canal es en realidad la diferencia entre dos términos de entropía que tienen una referencia en común, la información será la misma que la diferencia entre los términos de entropía diferencial correspondientes. Por lo tanto, disponemos de una justificación perfecta al usar el término $h(X)$, definido en la ecuación (9.66) como la entropía diferencial de la variable aleatoria continua X.

Cuando disponemos de un vector aleatorio continuo $\mathbf{X}$ compuesto por n variables aleatorias $X_1, X_2, \ldots, X_n$, definimos la entropía diferencial de $\mathbf{X}$ como la *integral n*

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_X(\mathbf{x}) \log_2\left[\frac{1}{f_X(\mathbf{x})}\right] d\mathbf{x} \tag{9.68}$$

donde $f_X(\mathbf{x})$ es la *función de densidad de probabilidad conjunta de* $\mathbf{X}$.

EJEMPLO 9.7 Distribución uniforme

Considere una variable aleatoria X distribuida uniformemente sobre el intervalo $(0, a)$. La función de densidad de probabilidad de X es

$$f_X(x) = \begin{cases} \frac{1}{a}, & 0 < x < a \\ 0, & \text{en otro caso} \end{cases}$$

Al aplicar la ecuación (9.66) a esta distribución, obtenemos

$$\begin{aligned} h(X) &= \int_0^a \frac{1}{a} \log(a)\, dx \\ &= \log a \end{aligned} \tag{9.69}$$

Advierta que $\log a < 0$ para $a < 1$. Así, este ejemplo muestra que, a diferencia de una variable aleatoria discreta, la entropía diferencial de una variable aleatoria continua puede ser negativa. ◀

EJEMPLO 9.8 Distribución gaussiana

Considere un par arbitrario de variables aleatorias X y Y, cuyas funciones de densidad de probabilidad se denotan respectivamente mediante $f_Y(x)$ y $f_X(x)$, donde x es meramente una variable simulada. Adaptando la desigualdad fundamental de la ecuación (9.12) a la situación presente, podemos escribir[9]

$$\int_{-\infty}^{\infty} f_Y(x) \log_2\left(\frac{f_X(x)}{f_Y(x)}\right) dx \le 0 \tag{9.70}$$

o, de modo equivalente,

$$-\int_{-\infty}^{\infty} f_Y(x) \log_2 f_Y(x)\, dx \le -\int_{-\infty}^{\infty} f_Y(x) \log_2 f_X(x)\, dx \tag{9.71}$$

La cantidad en el primer miembro de la ecuación (9.71) es la entropía diferencial de la variable aleatoria Y; de tal modo,

$$h(Y) \le -\int_{-\infty}^{\infty} f_Y(x) \log_2 f_X(x)\, dx \tag{9.72}$$

Suponga ahora que las variables aleatorias X y Y se describen de la manera siguiente:

- Las variables aleatorias X y Y tienen la *misma media* μ y la *misma varianza* σ^2.
- La variable aleatoria X presenta una *distribución gaussiana*, como se indica por medio de

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{9.73}$$

Por tanto, la sustitución de la ecuación (9.73) en la (9.72), y el cambio de la base del logaritmo de 2 a $e = 2.7183$, produce

$$h(Y) \le -\log_2 e \int_{-\infty}^{\infty} f_Y(x) \left(-\frac{(x-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)\right) dx \tag{9.74}$$

Después de esto reconocemos las propiedades de la variable aleatoria Y (dado que su media es μ y su varianza es σ^2):

$$\int_{-\infty}^{\infty} f_Y(x)\, dx = 1$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_Y(x)\, dx = \sigma^2$$

Es posible simplificar la ecuación (9.74) como

$$h(Y) \leq \tfrac{1}{2}\log_2(2\pi e\sigma^2) \tag{9.75}$$

La cantidad en el segundo miembro de la ecuación (9.75) es de hecho la entropía diferencial de la variable aleatoria gaussiana X:

$$h(X) = \tfrac{1}{2}\log_2(2\pi e\sigma^2) \tag{9.76}$$

Por último, al combinar las ecuaciones (9.75) y (9.76) podemos escribir

$$h(Y) \leq h(X), \quad \begin{cases} X\text{: variable aleatoria gaussiana} \\ Y\text{: otra variable aleatoria} \end{cases} \tag{9.77}$$

donde la igualdad se cumple si, y solamente si, Y = X.

Podemos resumir ahora los resultados de este importante ejemplo como las dos propiedades entrópicas de una variable aleatoria gaussiana:

1. Para una varianza finita σ^2, la variable aleatoria gaussiana tiene la entropía diferencial más grande alcanzable por cualquier variable aleatoria.
2. La entropía de una variable aleatoria gaussiana X está determinada únicamente por la varianza de X (es decir, es independiente de la media de X).

En realidad, a la propiedad 1 se debe que el modelo del canal gaussiano se use de manera tan amplia como un modelo conservativo en el estudio de sistemas de comunicación digitales. ◀

■ INFORMACIÓN MUTUA

Considere a continuación un par de variables aleatorias continuas X y Y. Por analogía con la ecuación (9.47), definimos la *información mutua* entre las variables aleatorias X y Y como sigue:

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left[\frac{f_X(x|y)}{f_X(x)}\right]dx\,dy \tag{9.78}$$

donde $f_{XY}(x, y)$ es la función de densidad de probabilidad conjunta de X y Y, y $f_X(x|y)$ es la función de densidad de probabilidad condicional de X, dado que Y = y. Además, por analogía con las ecuaciones (9.45), (9.50), (9.43) y (9.44) encontramos que la información mutua $I(X; Y)$ tiene las siguientes propiedades:

$$1.\ I(X;Y) = I(Y;X) \tag{9.79}$$

$$2.\ I(X;Y) \geq 0 \tag{9.80}$$

$$3.\ I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X) \tag{9.81}$$

El parámetro $h(X)$ es la entropía diferencial de X; lo mismo sucede para $h(Y)$. El parámetro $h(X|Y)$ es la *entropía diferencial condicional* de X, dada Y; éste se define mediante la doble integral [vea la ecuación (9.41)]

$$h(X|Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left[\frac{1}{f_X(x|y)}\right]dx\,dy \tag{9.82}$$

El parámetro $h(Y|X)$ es la entropía condicional de Y, dada X; se define de manera similar a $h(X|Y)$.

9.10 *Teorema de la capacidad de información*

En esta sección recurrimos a la idea de información mutua para formular el teorema de la capacidad de información correspondiente a *canales gaussianos limitado en potencia y limitado en banda.* Para ser más específicos, consideramos un proceso estacionario de media cero $X(t)$ que está limitado en banda a B hertz. Dejemos que X_k, $k = 1, 2, ..., K$, denote las variables aleatorias continuas que se obtienen por medio de un muestreo uniforme del proceso $X(t)$ a la velocidad de Nyquist de $2B$ muestras por segundo. Estas muestras se transmiten en T segundos por un canal ruidoso, también limitado en banda a B hertz. Por tanto, el número de muestras K está dado por

$$K = 2BT \tag{9.83}$$

Nos referimos a X_k como una muestra de la *señal transmitida.* La salida del canal está perturbada por *ruido blanco gaussiano aditivo* (AWGN) de media cero y densidad espectral de potencia $N_0/2$. El ruido está limitado en banda a B hertz. Considere que las variables aleatorias continuas Y_k, $k = 1, 2, ..., K$ denotan muestras de la señal recibida, según se indica mediante

$$Y_k = X_k + N_k, \quad k = 1, 2, \ldots, K \tag{9.84}$$

La muestra de ruido N_k es gaussiana con media cero y varianza dada por

$$\sigma^2 = N_0 B \tag{9.85}$$

Suponemos que las muestras Y_k, $k = 1, 2, ..., K$ son estadísticamente independientes.

Un canal para el cual el ruido y la señal recibida son como se describe en las ecuaciones (9.84) y (9.85) se conoce como *canal gaussiano sin memoria en tiempo discreto.* Éste se modela como en la figura 9.13. Sin embargo, para establecer enunciados significativos en torno al canal, tenemos que asignar un *costo* a cada entrada del canal. Por lo común, el transmisor es de *potencia limitada*, por lo que es razonable definir el costo como

$$E[X_k^2] = P, \quad k = 1, 2, \ldots, K \tag{9.86}$$

donde P es la *potencia transmitida promedio.* El *canal gaussiano limitado en potencia* descrito aquí no es sólo de importancia teórica sino también práctica en el sentido de que modela muchos canales de comunicación, incluso los enlaces de radio de línea de visión y de satélites.

La *capacidad de información* del canal se define como el máximo de la información mutua entre la entrada X_k del canal y la salida Y_k del canal sobre todas las distribuciones relativas a la entrada X_k que satisface la restricción de potencia de la ecuación (9.86). Considere que

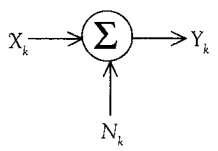


FIGURA 9.13 Modelo del canal gaussiano sin memoria en tiempo discreto.

$I(X_k; Y_k)$ denota la información mutua entre X_k y Y_k. Podemos definir entonces la capacidad de información del canal como

$$C = \max_{f_{X_k}(x)} \{I(X_k; Y_k) : E[X_k^2] = P\} \tag{9.87}$$

donde la maximización se efectúa con respecto a $f_{X_k}(x)$, la función de densidad de probabilidad de X_k.

La información mutua $I(X_k; Y_k)$ puede expresarse en una de las dos formas equivalentes que se muestran en la ecuación (9.81). Para el propósito presente, utilizamos la segunda línea de esta ecuación, por lo que escribimos

$$I(X_k; Y_k) = h(Y_k) - h(Y_k | X_k) \tag{9.88}$$

Puesto que X_k y N_k son variables aleatorias independientes, y su suma es igual a Y_k, como en la ecuación (9.84), encontramos que la entropía diferencial condicional de Y_k, dado X_k, es igual a la entropía diferencial de N_k (vea el problema 9.28):

$$h(Y_k | X_k) = h(N_k) \tag{9.89}$$

Por tanto, podemos reescribir la ecuación (9.88) en la forma

$$I(X_k; Y_k) = h(Y_k) - h(N_k) \tag{9.90}$$

Puesto que $h(N_k)$ es independiente de la distribución de X_k, la maximización de $I(X_k; Y_k)$ de acuerdo con la ecuación (9.87) requiere maximizar $h(Y_k)$, la entropía diferencial de la muestra Y_k de la señal recibida. Para que $h(Y_k)$ sea máxima, Y_k tiene que ser una variable aleatoria gaussiana (vea el ejemplo 9.8). Es decir, las muestras de la señal recibida representan un proceso semejante al ruido. En seguida observamos que si N_k se supone gaussiana, la muestra X_k de la señal transmitida también debe serlo. Por tanto, resulta viable establecer que la maximización especificada en la ecuación (9.87) se consigue eligiendo las muestras de la señal transmitida de un proceso semejante al ruido de potencia promedio P. En forma correspondiente, podemos reformular la ecuación (9.87) como

$$C = I(X_k; Y_k) : X_k \text{ gaussiana} \quad E[X_k^2] = P \tag{9.91}$$

donde la información mutua $I(X_k; Y_k)$ se define de acuerdo con la ecuación (9.90).

Con respecto a la evaluación de la capacidad de información C, procedemos en tres etapas:

1. La varianza de la muestra Y_k de la señal recibida es igual a $P + \sigma^2$. De ese modo, el uso de la ecuación (9.76) produce la entropía diferencial de Y_k como

$$h(Y_k) = \tfrac{1}{2} \log_2 [2\pi e(P + \sigma^2)] \tag{9.92}$$

2. La varianza de la muestra de ruido N_k es igual a σ^2. Por tanto, el empleo de la ecuación (9.76) genera la entropía diferencial de N_k como

$$h(N_k) = \tfrac{1}{2} \log_2 (2\pi e \sigma^2) \tag{9.93}$$

3. Sustituyendo las ecuaciones (9.92) y (9.93) en la (9.90) y admitiendo la definición de la capacidad dada en la ecuación (9.91), obtenemos el resultado deseado:

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2}\right) \text{ bits por transmisión} \tag{9.94}$$

Con el canal utilizado K veces para la transmisión de K muestras del proceso $X(t)$ en T segundos, encontramos que la capacidad de información por tiempo unitario es (K/T) veces el

resultado dado en la ecuación (9.94). El número K es igual a $2BT$, como en la ecuación (9.83). Por tanto, es posible expresar la capacidad de información en la forma equivalente:

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits por segundo} \tag{9.95}$$

donde hemos utilizado la ecuación (9.85) para la varianza del ruido σ^2.

Con base en la fórmula de la ecuación (9.95), es posible establecer ahora el tercero (y más famoso) teorema de Shannon, el *teorema de la capacidad de información*,[10] del modo siguiente:

La capacidad de información de un canal continuo de ancho de banda B hertz, perturbado por ruido blanco gaussiano aditivo de densidad espectral de potencia $N_0/2$ y limitado en ancho de banda a B, está dada por

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits por segundo}$$

donde P es la potencia promedio transmitida.

El teorema de la capacidad de información es uno de los resultados más notables de la teoría de la información, ya que, en una fórmula simple, subraya claramente la interrelación entre tres parámetros clave del sistema: el ancho de banda del canal, la potencia promedio transmitida (o, equivalentemente, la potencia promedio de la señal recibida) y la densidad espectral de potencia del ruido a la salida del canal. La dependencia de la capacidad de información C del ancho de banda del canal B es *lineal*, en tanto que su dependencia con respecto a la relación señal a ruido P/N_0B es *logarítmica*. Por tanto, *resulta más fácil aumentar la capacidad de información de un canal de comunicación al expandir su ancho de banda, que incrementar la potencia transmitida para una varianza de ruido preestablecida.*

El teorema implica que, para una potencia promedio transmitida P y ancho de banda de canal B dados, podemos transmitir información a una tasa de C bits por segundo, según se define en la ecuación (9.95), con una probabilidad de error arbitrariamente pequeña empleando sistemas de codificación suficientemente complejos. No es posible transmitir a una tasa superior que C bits por segundo mediante cualquier sistema de codificación sin una probabilidad definida de error. Por tanto, el teorema de la capacidad del canal define el *límite fundamental* en la tasa de transmisión sin errores para un canal gaussiano limitado en banda y limitado en potencia. Sin embargo, para atender este límite, la señal transmitida debe tener propiedades estadísticas que se aproximen a las del ruido blanco gaussiano.

Empaquetamiento de esferas[11]

Para proporcionar un argumento plausible que fundamente el teorema de capacidad de información, suponga que utilizamos un esquema de codificación que produce K palabras de código, una por cada muestra de la señal transmitida. Sea n la longitud (es decir, el número de bits) de cada palabra de código. Se presupone que el esquema de codificación está ideado para producir una probabilidad aceptablemente baja del error de símbolo. Además, las palabras de código satisfacen la restricción de potencia; es decir, la potencia promedio contenida en la transmisión de cada palabra de código con n bits es nP, donde P es la potencia promedio por bit.

Supóngase que se transmite cualquier palabra de código en el código. El vector recibido de n bits es de distribución gaussiana con media igual a la palabra de código transmitida y varianza igual a $n\sigma^2$, donde σ^2 es la varianza del ruido. Con alta probabilidad, el vector recibido se ubica dentro de una esfera de radio $\sqrt{n\sigma^2}$, centrada en la palabra de código transmitida. Dicha esfera se encuentra a su vez contenida en una gran esfera de radio $\sqrt{n(P+\sigma^2)}$, donde $n(P + \sigma^2)$ es la potencia promedio del vector recibido.

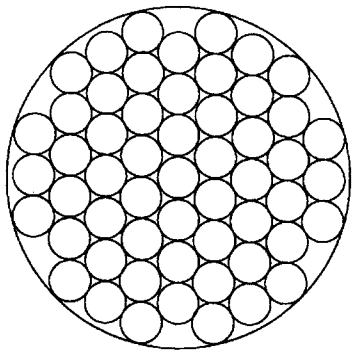


FIGURA 9.14 El problema del empaquetamiento de esferas.

De esta manera es posible visualizar la imagen que se presenta en la figura 9.14. Con todo lo que se encuentra dentro de una pequeña esfera de radio $\sqrt{n\sigma^2}$ asignado a la palabra de código sobre la cual se centra, resulta razonable afirmar que cuando se transmite esta palabra de código particular, es alta la probabilidad de que el vector recibido se encontrará dentro de la esfera de "decodificación" correcta. La pregunta clave es: ¿cuántas esferas de decodificación pueden empacarse dentro de la esfera mayor de vectores recibidos? En otros términos, ¿cuántas palabras de código podemos realmente elegir? Para responder a esta pregunta, reconocemos primero que el volumen de una esfera n dimensional de radio r puede escribirse como $A_n r^n$, donde A_n es un factor de escalamiento. Por tanto, podemos plantear los siguientes enunciados:

- El volumen de una esfera de vectores recibidos es $A_n[n(P + \sigma^2)]^{n/2}$.
- El volumen de la esfera de decodificación es $A_n(n\sigma^2)^{n/2}$.

Por lo tanto, se concluye que el máximo de esferas de decodificación *que no se interceptan* y que pueden estar empacadas dentro de la esfera de posibles vectores recibidos es

$$\frac{A_n[n(P+\sigma^2)]^{n/2}}{A_n(n\sigma^2)^{n/2}} = \left(1+\frac{P}{\sigma^2}\right)^{n/2} = 2^{(n/2)\log_2(1+P/\sigma^2)} \tag{9.96}$$

Tomando el logaritmo de este resultado con respecto a la base 2, vemos de inmediato que el número máximo de bits por transmisión para una baja probabilidad de error es en realidad como se definió antes en la ecuación (9.94).

EJEMPLO 9.9 Reconfiguración de la constelación para potencia reducida

Para ilustrar la idea de empaquetamiento de esferas, considere la constelación cuadrada QAM-64 de la figura 9.15*a*. Esta figura describe dos esferas de decodificación bidimensionales que no se interceptan centradas en los puntos de mensaje en la constelación. Al tratar de empacar lo más apretadamente posible las esferas de codificación mientras se mantiene la misma distancia euclidiana entre los puntos de mensaje que antes, obtenemos la constelación alternativa que se ilustra en la figura 9.15*b*. Con una distancia euclidiana en común entre los puntos de mensaje, las dos constelaciones de la figura 9.15 producen aproximadamente la misma tasa de errores de bit, suponiendo el uso de una relación señal a ruido suficientemente elevada con respecto a un canal AWGN; vea, por ejemplo, la ecuación (5.95). Sin embargo, comparando estas dos constelaciones, encontramos que la suma de las distancias euclidianas al cuadrado desde los puntos de mensaje hasta el origen en la figura 9.15*b* es más pequeña que en la figura 9.15*a*. Por tanto, se concluye que la constelación más apretada de la figura 9.15*b* tiene

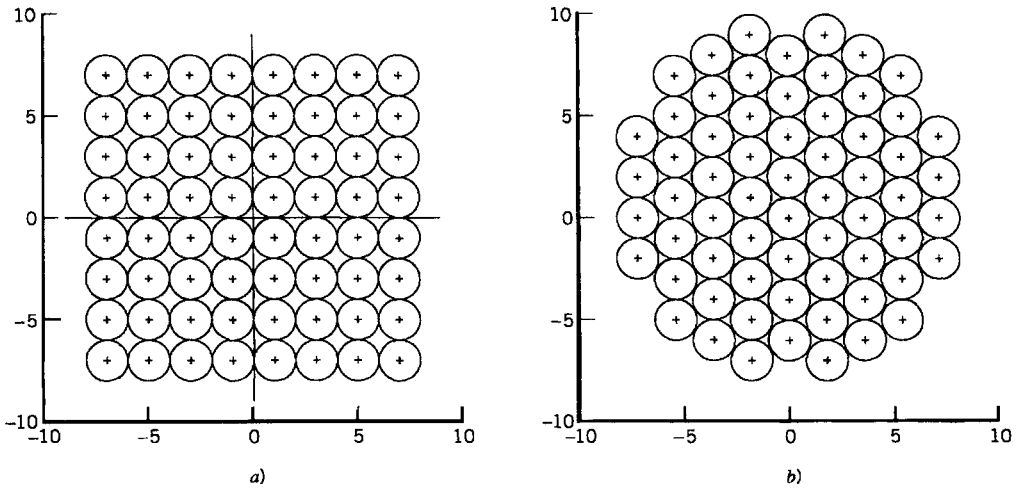


FIGURA 9.15 *a*) Constelación cuadrada QAM 64. *b*) La alternativa de vinculación más apretada para la de la parte *a*.

una ventaja sobre la constelación cuadrada de la figura 9.15*a*: una energía por símbolo de señal promedio transmitida más pequeña para la misma tasa de error de bits sobre un canal AWGN. ◀

9.11 *Implicaciones del teorema de la capacidad de información*

Ahora que tenemos una percepción intuitiva del teorema de la capacidad de información, podemos proseguir para explicar sus implicaciones en el contexto de un canal gaussiano que está limitado tanto en potencia como en ancho de banda. Sin embargo, para que la exposición sea de utilidad, necesitamos un marco de referencia ideal contra el cual pueda valorarse el desempeño de un sistema de comunicación práctico. Con ese fin introducimos la noción de un *sistema ideal* definido como uno que transmite datos a una tasa de bits R_b igual a la capacidad de información C. Es posible expresar entonces la potencia promedio transmitida como

$$P = E_b C \tag{9.97}$$

donde E_b es la energía transmitida por bit. Por tanto, el sistema ideal se define mediante la ecuación

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{N_0}\frac{C}{B}\right) \tag{9.98}$$

De manera equivalente, podemos definir la relación de *la energía de la señal por bit a la densidad espectral de la potencia del ruido* E_b/N_0 en términos de la razón C/B para el sistema ideal como

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{CB} \tag{9.99}$$

Una gráfica de la eficiencia del ancho de banda R_b/B en función de E_b/N_0 recibe el nombre de *diagrama de eficiencia del ancho de banda*. Una forma genérica de este diagrama se ilustra en la

figura 9.16, donde la curva denominada "frontera de la capacidad" corresponde al sistema ideal para el cual $R_b = C$. Con base en la figura 9.16, es factible señalar las siguientes observaciones:

1. Para *ancho de banda infinito*, la razón E_b/N_0 tiende al valor límite

$$\left(\frac{E_b}{N_0}\right)_\infty = \lim_{B\to\infty}\left(\frac{E_b}{N_0}\right) = \log 2 = 0.693 \tag{9.100}$$

Este valor se denomina *límite de Shannon* para un canal AWGN, suponiendo una tasa de código de cero. Expresado en decibeles, es igual a −1.6 dB. El valor límite correspondiente de la capacidad del canal se obtiene dejando que el ancho de banda B del canal en la ecuación (9.95) tienda a infinito; consecuentemente encontramos que

$$C_\infty = \lim_{B\to\infty} C = \frac{P}{N_0}\log_2 e \tag{9.101}$$

donde e es la base del logaritmo natural.

2. La *frontera de la capacidad*, definida por la curva para la tasa de bits crítica $R_b = C$, separa las combinaciones de parámetros del sistema que tienen el potencial para soportar una transmisión sin errores ($R_b < C$) de aquellas para las cuales no es posible ese tipo de transmisión ($R_b > C$). La última región se muestra sombreada en la figura 9.16.
3. El diagrama subraya los *compromisos* potenciales entre E_b/N_0, R_b/B y la probabilidad del error de símbolo P_e. En particular, podemos observar el movimiento del punto de operación a lo

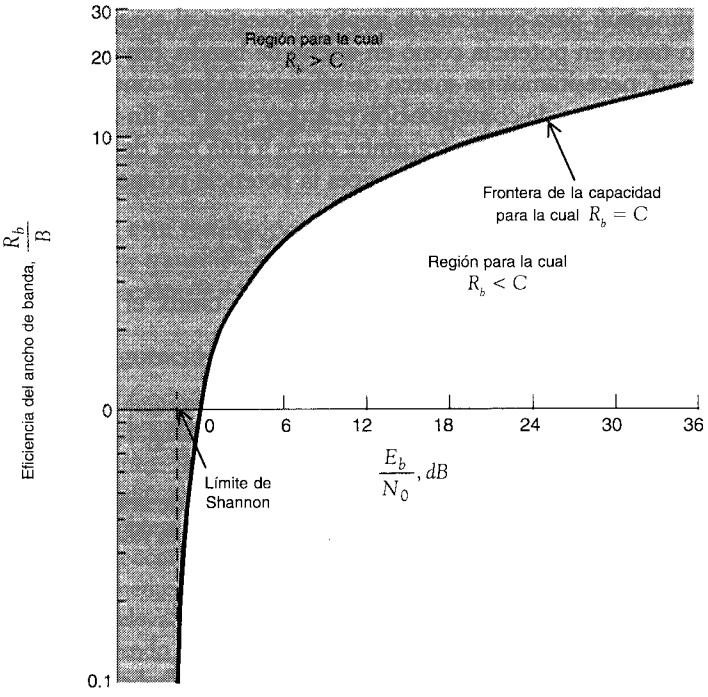


FIGURA 9.16 Diagrama de eficiencia del ancho de banda.

largo de una línea horizontal como el intercambio de P_e en función de E_b/N_0 para una R_b/B fija. Por otra parte, es posible advertir el movimiento del punto de operación a lo largo de una línea vertical como el intercambio de P_e en función de R_b/B para una E_b/N_0 fija.

EJEMPLO 9.10 PCM M-aria

En este ejemplo consideramos un sistema PCM M-ario a la luz del teorema de la capacidad del canal bajo la suposición de que el sistema opera por arriba del umbral de error. Es decir, la probabilidad promedio de error debido al ruido del canal es despreciable.

Suponemos que el sistema PCM M-ario utiliza una palabra de código consistente en n elementos de código, teniendo cada uno M niveles de amplitud discretos posibles; de ahí el nombre "M-ario". Del capítulo 3 recordamos que para que un sistema PCM opere por arriba del umbral de error, debe proporcionarse un margen de ruido que sea lo suficientemente grande para mantener una tasa de error ignorable debida al ruido del canal. Esto, a su vez, significa que debe haber cierta separación entre estos M niveles de amplitud discretos. Denominando a esta separación $k\sigma$, donde k es una constante y $\sigma^2 = N_0B$ es la varianza de ruido medida en un ancho de banda de canal B. El número de niveles de amplitud M suele ser una potencia entera de 2. La potencia transmitida promedio será mínima si el intervalo de amplitud es simétrico en torno a cero. En ese caso, los niveles de amplitud discretos, normalizados con respecto a la separación $k\sigma$, tendrán los valores $\pm 1/2$, $+3/2$, ..., $\pm(M-1)/2$. Suponemos que estos M niveles de amplitud diferentes son igualmente probables. En consecuencia, encontramos que la potencia transmitida promedio está dada por

$$\begin{aligned} P &= \frac{2}{M}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \ldots + \left(\frac{M-1}{2}\right)^2\right](k\sigma)^2 \\ &= k^2\sigma^2\left(\frac{M^2-1}{12}\right) \end{aligned} \tag{9.102}$$

Suponga que el sistema PCM M-ario descrito aquí se utiliza para transmitir una señal de mensaje con su componente de frecuencia más alta igual a W hertz. La señal se muestrea a la frecuencia de Nyquist de $2W$ muestras por segundo. Suponemos que el sistema utiliza un cuantizador del tipo de ascenso medio, con L niveles de representación igualmente probables. Por tanto, la probabilidad de ocurrencia de cualquiera de los L niveles de representación es $1/L$. De manera correspondiente, la cantidad de información que transporta una muestra simple de una señal es $\log_2 L$ bits. Con una frecuencia de muestreo máxima de $2W$ muestras por segundo, la velocidad máxima de transmisión de información del sistema PCM medida en bits por segundo está dada por

$$R_b = 2W \log_2 L \text{ bits por segundo} \tag{9.103}$$

Puesto que el sistema PCM utiliza una palabra de código consistente en n elementos de código, teniendo cada uno M posibles valores de amplitud discretos, tenemos M^n diferentes palabras de código posibles. En un proceso de codificación único requerimos

$$L = M^n \tag{9.104}$$

Claramente, la velocidad de transmisión de información en el sistema no se afecta por el uso de un proceso de codificación. Por tanto, podemos eliminar L entre las ecuaciones (9.103) y (9.104) para obtener

$$R_b = 2Wn \log_2 M \text{ bits por segundo} \tag{9.105}$$

La ecuación (9.102) define la potencia transmitida promedio que se requiere para mantener un sistema PCM M-ario operando por arriba del umbral de error. De ese modo, al resolver esta ecuación para el número de niveles de amplitud discretos M, se obtiene

$$M = \left(1 + \frac{12P}{k^2N_0B}\right)^{1/2} \tag{9.106}$$

donde $\sigma^2 = N_0B$ es la varianza del ruido del canal medida en el ancho de banda B. Por tanto, sustituyendo la ecuación (9.106) en la (9.105), obtenemos

$$R_b = Wn \log_2\left(1+\frac{12P}{k^2N_0B}\right) \tag{9.107}$$

El ancho de banda del canal B requerido para transmitir un pulso rectangular de duración $1/2nW$ (que representa a un elemento de código en la palabra de código) está dado por (vea el capítulo 3)

$$B = \kappa n W$$

donde κ es una constante con un valor que se encuentra entre 1 y 2. Utilizando el valor mínimo posible $\kappa = 1$, encontramos que el ancho de banda del canal $B = nW$. De ese modo es posible reescribir la ecuación (9.107) como

$$R_b = B \log_2\left(1+\frac{12P}{k^2N_0B}\right) \tag{9.108}$$

El *sistema ideal* se describe mediante el teorema de Shannon de la capacidad del canal, dado en la ecuación (9.95). De esa manera, al comparar la ecuación (9.108) con la (9.95), advertimos que éstas son idénticas si la potencia transmitida promedio en el sistema PCM se incrementa en un factor $k^2/12$, en comparación con el sistema ideal. Quizás el punto más interesante que señalar en relación con la ecuación (9.108) sea que la forma de esta misma es correcta: *la potencia y el ancho de banda en un sistema PCM se intercambian de acuerdo con una base logarítmica, y la capacidad de información C es proporcional al ancho de banda B del canal.* ◀

▶ EJEMPLO 9.11 PSK M-aria y FSK M-aria

En este ejemplo comparamos las capacidades de intercambio de ancho de banda-potencia de señales PSK M-arias y FSK M-arias a la luz del teorema de Shannon de la capacidad de información. Considere primero un sistema PSK M-ario coherente que emplea un conjunto *no ortogonal* de M señales corridas en fase para la transmisión de datos binarios. Cada señal en el conjunto representa un símbolo con M $\log_2$ bits. Utilizando la definición de ancho de banda de nulo a nulo, es posible expresar la eficiencia del ancho de banda de la PSK M-aria como sigue [vea la ecuación (6.51)]:

$$\frac{R_b}{B} = \frac{\log_2 M}{2}$$

En la figura 9.17*a* mostramos los puntos de operación para números diferentes de niveles de fase M = 2, 4, 8, 16, 32, 64. Cada punto corresponde a una probabilidad del error de símbolo $P_e = 10^{-5}$. En la figura hemos incluido también la frontera de capacidad para el sistema ideal. Observamos en la figura 9.17*a* que cuando aumenta M, se mejora la eficiencia del ancho de banda, aunque el valor de E_b/N_0 requerido para la transmisión sin errores se aparta del límite de Shannon.

Considere a continuación un sistema FSK M-ario que utiliza un conjunto *ortogonal* de M señales corridas en frecuencia para la transmisión de datos binarios, con la separación entre frecuencias de señal adyacentes fijada en $1/2T$, donde T es el periodo del símbolo. Como con la PSK M-aria, cada señal en el conjunto representa un símbolo con $\log_2$ M bits. La eficiencia del ancho de banda de la FSK M-aria es como sigue [vea la ecuación (6.143)]:

$$\frac{R_b}{B} = \frac{2 \log_2 M}{M}$$

En la figura 9.17*b* mostramos los puntos de operación para números diferentes de niveles de frecuencia M = 2, 4, 8, 16, 32, 64 relativos a una probabilidad promedio del error de símbolo $P_e = 10^{-5}$. En la figura hemos incluido también la frontera de capacidad para el sistema ideal. Observamos que el incremento de M en la FSK M-aria (ortogonal) tiene el efecto opuesto al que se presenta en la PSK M-aria (no ortogonal). En particular, cuando M aumenta, lo que equivale a un requerimiento incrementado del ancho de banda, el punto de operación se acerca al límite de Shannon. ◀

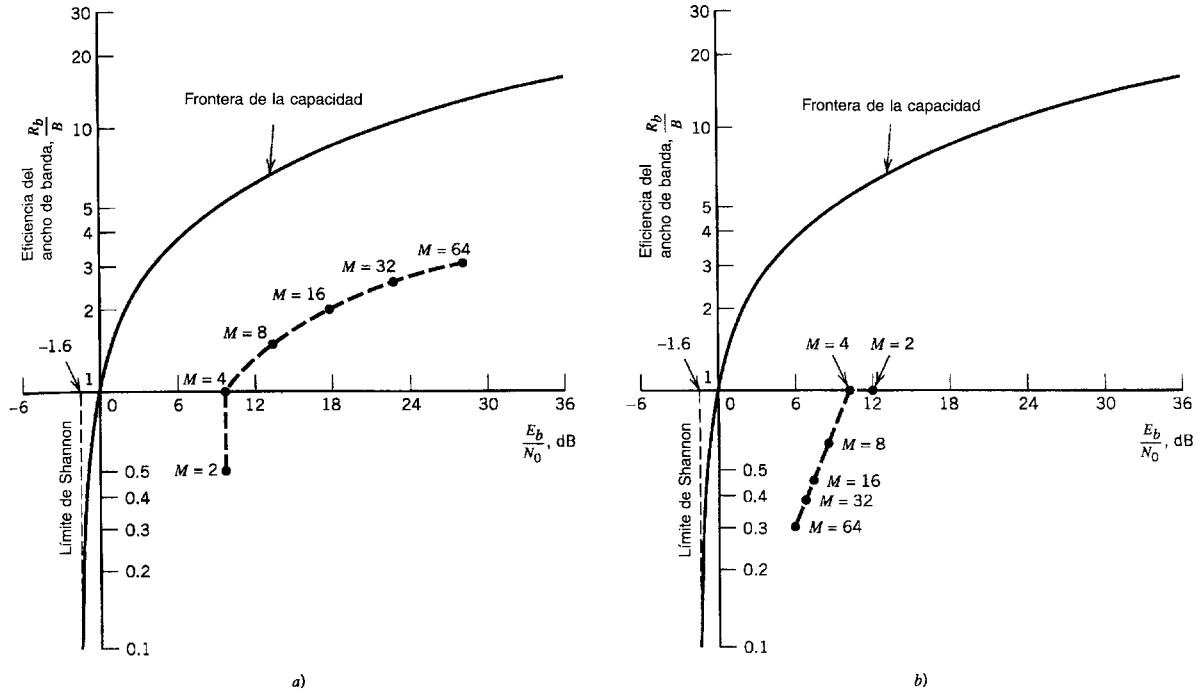


FIGURA 9.17 *a*) Comparación de la PSK M-aria contra el sistema ideal para $P_e = 10^{-5}$ y M creciente. *b*) Comparación de la FSK M-aria contra el sistema ideal para $P_e = 10^{-5}$ y M creciente.

EJEMPLO 9.12 Capacidad del canal AWGN de entrada binaria

En este ejemplo investigamos la capacidad de un canal AWGN utilizando transmisión de señales antípodas binarias *codificadas* (es decir, niveles -1 y $+1$ para símbolos binarios 0 y 1, respectivamente). En particular, consideramos el aspecto de determinar la tasa de error de bits alcanzable mínima como una función de E_b/N_0 para tasa de código variable r. Se supone que los símbolos binarios 0 y 1 son igualmente probables.

Dejemos que las variables aleatorias X y Y denoten la entrada y la salida del canal, respectivamente; X es una variable discreta, en tanto que Y es una continua. A la luz de la segunda línea de la ecuación (9.81), podemos expresar la información mutua entre la entrada y la salida del canal como

$$I(X;Y) = h(Y) - h(Y|X)$$

el segundo término, $h(Y|X)$, es la entropía diferencial condicional de la salida Y del canal, dada la entrada X del canal. En virtud de las ecuaciones (9.89) y (9.93), este término es exactamente la entropía de la distribución gaussiana. Por tanto, al utilizar σ^2 para denotar la varianza del ruido del canal, es posible escribir

$$h(Y|X) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Después de esto, el primer término, $h(Y)$, es la entropía diferencial de la salida del canal Y. Con el uso de transmisión de señales antípodas binarias, la función de densidad de probabilidad de Y, dada $X = x$, es una mezcla de dos distribuciones gaussianas con varianza en común σ^2 y valores medios -1 y $+1$, según indica

$$f_Y(y_i|x) = \frac{1}{2}\left[\frac{\exp(-(y_i+1)^2/2\sigma^2)}{\sqrt{2\pi}\sigma} + \frac{\exp(-(y_i-1)^2/2\sigma^2)}{\sqrt{2\pi}\sigma}\right] \tag{9.109}$$

Por tanto, podemos determinar la entropía diferencial de Y utilizando la fórmula

$$h(Y) = -\int_{-\infty}^{\infty} f_Y(y_i|x) \log_2[f_Y(y_i|x)]\, dy_i$$

donde $f_Y(y_i|x)$ se define mediante la ecuación (9.109). A partir de las fórmulas de $h(Y|X)$ y $h(Y)$, resulta claro que la información mutua es solamente una función de la varianza del ruido σ^2. Utilizando $M(\sigma^2)$ para denotar esta dependencia funcional, es posible escribir entonces

$$I(X;Y) = M(\sigma^2)$$

Desafortunadamente, no hay una fórmula cerrada que podamos obtener para $M(\sigma^2)$ debido a la dificultad de determinar $h(Y)$. A pesar de eso, la entropía diferencial $h(Y)$ se puede aproximar bastante bien utilizando *integración de Monte Carlo*, la cual se programa directamente en una computadora digital; vea el problema 9.36.

Debido a que los símbolos 0 y 1 son igualmente probables, se concluye que la capacidad C del canal es igual a la información mutua entre X y Y. Por tanto, para la transmisión de datos sin errores por el canal AWGN, la tasa de código r debe satisfacer la condición

$$r < M(\sigma^2) \tag{9.110}$$

Una medida robusta de la razón E_b/N_0 es

$$\frac{E_b}{N_0} = \frac{P}{N_0 r} = \frac{P}{2\sigma^2 r}$$

donde P es la potencia transmitida promedio, y $N_0/2$ es la densidad espectral de potencia bilateral del ruido del canal. Sin pérdida de generalidad, podemos dejar $P = 1$. En consecuencia, resulta factible expresar la varianza del ruido como

$$\sigma^2 = \frac{N_0}{2E_b r} \tag{9.111}$$

Al sustituir la ecuación (9.111) en la (9.110) y reordenar términos, tenemos la relación que se desea:

$$\frac{E_b}{N_0} = \frac{1}{2rM^{-1}(r)} \tag{9.112}$$

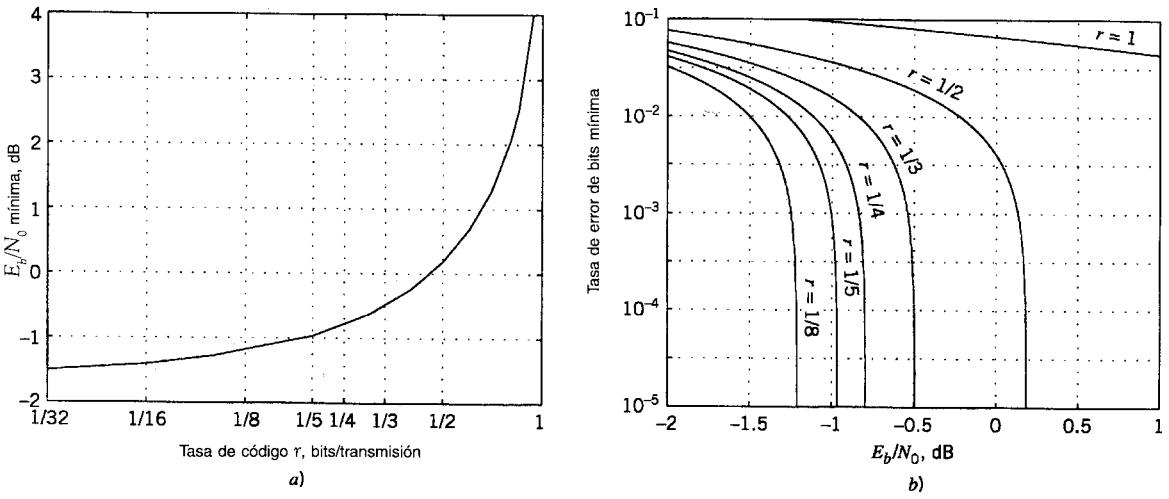


FIGURA 9.18 Transmisión de señales antipodales binarias por un canal AWGN. *a*) E_b/N_0 mínima en función de la tasa de código r. *b*) Tasa de error de bits (BER) mínima en función de E_b/N_0 para tasa de código r variable.

donde $M^{-1}(r)$ es el *inverso* de la información mutua entre la entrada y la salida del canal, expresado como una función de la tasa de código r.

Recurriendo al método de Monte Carlo para estimar la entropía diferencial $h(Y)$ y, en consecuencia, $M^{-1}(r)$, se calculan las gráficas de la figura 9.18.[12] En la figura 9.18*a* se grafica el mínimo E_b/N_0 en función de la tasa de código r para comunicación sin errores. En la figura 9.18*b* se grafica la tasa de error de bits mínima alcanzable en función de E_b/N_0 con la tasa de código r como un parámetro ejecutable. A partir de la figura 9.18 es posible extraer las siguientes conclusiones:

- Para una transmisión de señales no codificada (es decir, $r = 1$), se requiere una E_b/N_0 infinita para la comunicación sin errores, lo cual concuerda con lo que sabemos acerca de la transmisión de datos no codificados por un canal AWGN.
- La E_b/N_0 mínima disminuye con la reducción de la tasa de código r, lo cual se satisface de manera intuitiva. Por ejemplo, para $r = 1/2$, el valor mínimo de E_b/N_0 es un poco menor que 0.1 dB.
- Cuando r tiende a cero, la E_b/N_0 mínima tiende al valor límite de -1.6 dB, lo que concuerda con el límite de Shannon obtenido antes; vea la ecuación (9.100). ◀

9.12 *Capacidad de información del canal de ruido coloreado*[13]

El teorema de la capacidad de información como se formula en la ecuación (9.95) se aplica a un canal con ruido blanco limitado en banda. En esta sección extendemos el teorema de Shannon de la capacidad de información a un caso más general de un *canal ruidoso no blanco* o *coloreado*. Específicamente, considere el modelo de canal que se muestra en la figura 9.19*a* donde la función de transferencia de este mismo se denota por medio de $H(f)$. El ruido del canal $n(t)$, el cual aparece aditivamente a la salida del canal, se modela como la función de muestreo de un proceso gaussiano estacionario de media cero y densidad espectral de potencia $S_N(f)$. El requerimiento es doble:

1. Determinar el conjunto de entrada, descrito por la densidad espectral de potencia $S_X(f)$, que maximice la información mutua entre la salida del canal $y(t)$ y la entrada $x(t)$ del mismo, sujeta a la restricción de que la potencia promedio de $x(t)$ se fija en un valor constante P.
2. Por tanto, determinar la capacidad de información óptima del canal.

Éste es un problema de optimización restringida y para resolverlo procedemos de la forma que sigue:

- Debido a que el canal es lineal, es posible sustituir el modelo de la figura 9.19*a* con el modelo equivalente que se presenta en la figura 9.19*b*. Desde el punto de vista de las características espectrales de la señal más el ruido medido a la salida del canal, los dos modelos de la figura 9.19 resultan equivalentes, siempre y cuando la densidad espectral de potencia del ruido

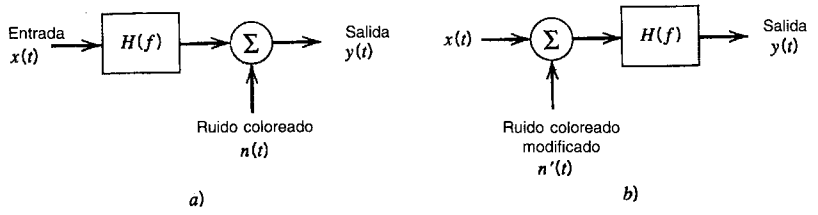


FIGURA 9.19 *a*) Modelo limitado en banda, canal ruidoso limitado en potencia. *b*) Modelo equivalente del canal.

$n'(t)$ en la figura 9.19*b* se defina en términos de la densidad espectral de potencia del ruido $n(t)$ en la figura 9.19*a* como

$$S_{N'}(f) = \frac{S_N(f)}{|H(f)|^2} \tag{9.113}$$

donde $|H(f)|$ es la respuesta en magnitud en el canal.

- Para simplificar el análisis utilizamos el "principio de divide y vencerás", de una manera similar a la descrita en la sección 6.12. De manera específica, el canal se divide en un número grande de franjas de frecuencia adyacentes, como se ilustra en la figura 9.20. Cuanto más pequeño hacemos el intervalo de frecuencia incremental Δf de cada subcanal, tanto mejor es la aproximación.

El resultado neto de estos dos puntos es que el modelo original de la figura 9.19*a* se sustituye por la combinación en paralelo de un número finito de subcanales, N, cada uno de los cuales se corrompe esencialmente por un "ruido blanco gaussiano limitado en banda".

El subcanal k-ésimo en la aproximación al modelo de la figura 9.19*b* se describe por medio de

$$y_k(t) = x_k(t) + n_k(t), \quad k = 1, 2, \ldots, N \tag{9.114}$$

La potencia promedio de la componente de señal $x_k(t)$ es

$$P_k = S_X(f_k)\,\Delta f, k = 1, 2, \ldots, N \tag{9.115}$$

donde $S_X(f_k)$ es la densidad espectral de potencia de la señal de entrada evaluada a la frecuencia $f = f_k$. La varianza de la componente de ruido $n_k(t)$ es

$$\sigma_k^2 = \frac{S_N(f_k)}{|H(f_k)|^2}\Delta f, \quad k = 1, 2, \ldots, N \tag{9.116}$$

donde $S_N(f_k)$ y $|H(f_k)|$ son la densidad espectral de ruido y la respuesta en magnitud de canal evaluadas a la frecuencia f_k, respectivamente. La capacidad de información del k-ésimo canal es

$$C_k = \frac{1}{2}\Delta f \log_2\left(1 + \frac{P_k}{\sigma_k^2}\right), \quad k = 1, 2, \ldots, N \tag{9.117}$$

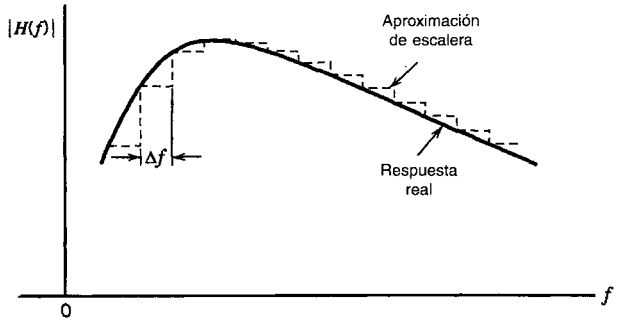


FIGURA 9.20 Aproximación de escalera de una respuesta en magnitud arbitraria $|H(f)|$; sólo se muestra la porción de frecuencia positiva de la respuesta.

donde el factor 1/2 toma en cuenta el hecho de que Δf se aplica tanto a las frecuencias positivas como negativas. Todos los N subcanales son independientes entre sí. Por ello la capacidad total del canal completo está dada aproximadamente por la sumatoria

$$\begin{aligned} C &\simeq \sum_{k=1}^{N} C_k \\ &= \frac{1}{2}\sum_{k=1}^{N} \Delta f \log_2\left(1+\frac{P_k}{\sigma_k^2}\right) \end{aligned} \tag{9.118}$$

El problema que debemos abordar es maximizar la capacidad de información total C sujeta a la restricción

$$\sum_{k=1}^{N} P_k = P = \text{constante} \tag{9.119}$$

El procedimiento usual para resolver un problema de optimización restringida reside en utilizar el *método de los multiplicadores de Lagrange;* vea la nota 19 en el capítulo 6. Para proceder con esta optimización, definimos primero una función objetivo que incorpora tanto la capacidad de información C como la restricción [es decir, las ecuaciones (9.118) y (9.119)], como se indica mediante

$$J = \frac{1}{2}\sum_{k=1}^{N} \Delta f \log_2\left(1+\frac{P_k}{\sigma_k^2}\right) + \lambda\left(P - \sum_{k=1}^{N} P_k\right) \tag{9.120}$$

donde λ es el multiplicador de Lagrange. Después de esto, al diferenciar la función objetivo J con respecto a P_k e igualar a cero el resultado, obtenemos

$$\frac{\Delta f \log_2 e}{P_k + \sigma_k^2} - \lambda = 0$$

Para satisfacer esta solución de optimización, imponemos el siguiente requerimiento:

$$P_k + \sigma_k^2 = K\,\Delta f \quad \text{para } k = 1, 2, \ldots, N \tag{9.121}$$

donde K es una constante igual para toda k. La constante K se elige para satisfacer la restricción de potencia promedio.

Al insertar los valores de definiciones de las ecuaciones (9.115) y (9.116) en la condición de optimización de la ecuación (9.121), simplificando y reordenando términos, se llega a

$$S_X(f_k) = K - \frac{S_N(f_k)}{|H(f_k)|^2}, \quad k = 1, 2, \ldots, N \tag{9.122}$$

Sea $\mathscr{F}_A$ el intervalo de frecuencia para el cual la constante K satisface la condición

$$K \geq \frac{S_N(f)}{|H(f)|^2}$$

Entonces, cuando se deja que el intervalo de frecuencia incremental Δf tienda a cero y el número de subcanales N se aproxime a infinito, es posible utilizar la ecuación (9.122) para establecer formalmente que la densidad espectral de potencia del conjunto de entrada que alcanza la capacidad de información óptima es una cantidad no negativa definida por

$$S_X(f) = \begin{cases} K - \dfrac{S_N(f)}{|H(f)|^2} & \text{para } f \in \mathscr{F}_A \\ 0 & \text{en otro caso} \end{cases} \tag{9.123}$$

Puesto que la potencia promedio de un proceso aleatorio es el área total bajo la curva de la densidad espectral de potencia del proceso, podemos expresar la potencia promedio de la entrada $x(t)$ como

$$P = \int_{f \in \mathscr{F}_A} \left(K - \frac{S_N(f)}{|H(f)|^2} \right) df \tag{9.124}$$

Para una P preestablecida y $S_N(f)$ y $H(f)$ especificadas, la constante K es la solución de la ecuación (9.124).

El único aspecto cuya realización nos queda pendiente es determinar la capacidad de información óptima. Sustituyendo la solución de optimización de la ecuación (9.121) en la (9.118) y utilizando después los valores de definición de las ecuaciones (9.115) y (9.116), obtenemos

$$C \simeq \frac{1}{2} \sum_{k=1}^{N} \Delta f \log_2 \left(K \frac{|H(f_k)|^2}{S_N(f_k)} \right)$$

Cuando se deja que tienda a cero el intervalo de frecuencia incremental Δf, esta ecuación toma la forma límite:

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \log_2 \left(K \frac{|H(f)|^2}{S_N(f)} \right) df \tag{9.125}$$

donde la constante K se elige como la solución de la ecuación (9.124) para una señal de potencia de entrada P preestablecida.

■ INTERPRETACIÓN DE LA ANALOGÍA DEL LLENADO DE AGUA DEL TEOREMA DE LA CAPACIDAD DE INFORMACIÓN

Las ecuaciones (9.123) y (9.124) sugieren la imagen que se ilustra en la figura 9.21. Específicamente, hacemos las siguientes observaciones:

- La densidad espectral de potencia de entrada $S_X(f)$ apropiada se describe como las regiones inferiores de la función $S_N(f)/|H(f)|^2$ que se encuentra por debajo del nivel constante K, las cuales se presentan sombreadas.
- La potencia P de entrada se define por medio del área total de esas regiones sombreadas.

La imagen del dominio espectral que se ilustra aquí recibe el nombre de *interpretación del llenado de agua (vertido)*, en el sentido de que el proceso por medio del cual la potencia de entrada se

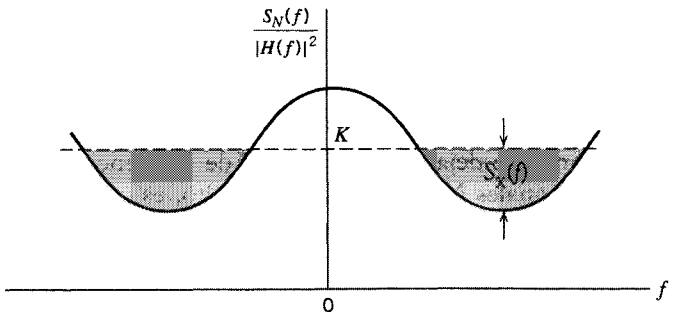


FIGURA 9.21 Interpretación de la analogía del llenado de agua relativa al teorema de capacidad de información para un canal con ruido coloreado.

distribuye a través de la función $S_N(f)/|H(f)|^2$ es idéntico a la forma en la cual el agua misma se distribuye en un recipiente.

Considere ahora el caso idealizado de una señal limitada en banda de ruido blanco gaussiano aditivo de densidad espectral de potencia $N(f) = N_0/2$. La función de transferencia $H(f)$ es la de un filtro pasobanda definida por

$$H(f) = \begin{cases} 1, & 0 \le f_c - \dfrac{B}{2} \le |f| \le f_c + \dfrac{B}{2} \\ 0, & \text{en otro caso} \end{cases}$$

donde f_c es la frecuencia de banda media y B es el ancho de banda del canal. En este caso especial, las ecuaciones (9.124) y (9.125) se reducen, respectivamente, a

$$P = 2B\left(K - \frac{N_0}{2}\right)$$

$$C = B\log_2\left(\frac{2K}{N_0}\right)$$

Por tanto, al eliminar K entre estas dos ecuaciones, obtenemos la forma estándar del teorema de capacidad de Shannon, definido por la ecuación (9.95).

▶ Ejemplo 9.13 Capacidad del canal dominado por la NEXT

A partir de la exposición que se presentó en la sección 4.8, recordamos que uno de los principales deterioros del canal en las líneas de suscriptor digitales es la diafonía del extremo cercano (NEXT). La densidad espectral de potencia de esta diafonía puede considerarse como

$$S_N(f) = |H_{\text{NEXT}}(f)|^2 \, S_X(f) \tag{9.126}$$

donde $S_X(f)$ es la densidad espectral de potencia de la señal transmitida y $H_{\text{NEXT}}(f)$ es la función de transferencia que acopla los pares trenzados adyacentes. La única restricción que tenemos que satisfacer en este ejemplo es que la función de densidad espectral de potencia $S_X(f)$ sea *no negativa para toda* f. Sustituyendo la ecuación (9.126) en la (9.123), encontramos de inmediato que esta condición se cumple al resolver K como

$$K = \left(1 + \frac{|H_{\text{NEXT}}(f)|^2}{|H(f)|^2}\right) S_X(f)$$

Por último, utilizando este resultado en la ecuación (9.125), encontramos que la capacidad de canal de suscriptor digital dominada por la NEXT está dada por

$$C = \frac{1}{2}\int_{\mathcal{F}_A} \log_2\left(1 + \frac{|H(f)|^2}{|H_{\text{NEXT}}(f)|^2}\right) df$$

donde $\mathcal{F}_A$ es el conjunto de frecuencias positivas y negativas para las cuales $S_X(f) > 0$. ◀

9.13 *Teoría de distorsión de velocidad*

En la sección 9.3 presentamos el teorema de la codificación de fuente para una fuente discreta sin memoria, de acuerdo con el cual la longitud promedio de palabra de código debe ser al menos tan grande como la entropía de la fuente en la codificación perfecta (es decir, representación perfecta de

la fuente). Sin embargo, en muchas situaciones prácticas existen restricciones que fuerzan a que la codificación resulte imperfecta, de lo cual se desprende una *distorsión* inevitable. Por ejemplo, las restricciones impuestas por un canal de comunicación pueden imponer un límite superior en la tasa de código permisible y, por tanto, en la longitud de palabra de código asignada a la fuente de información. Como otro ejemplo, la fuente de información quizá tenga una amplitud constante, como en el caso del habla, y el requerimiento sea cuantizar la amplitud de cada muestra generada por la fuente para permitir su representación por medio de una palabra de código de longitud finita, como la modulación por codificación de pulsos. En estos casos, el problema se conoce como la *codificación de fuente con un criterio de fidelidad*, y la rama de la teoría de la información que lo aborda se conoce como *teoría de distorsión de velocidad*,[14] la cual encuentra aplicaciones en dos tipos de situaciones:

- Codificación de fuente donde el alfabeto de codificación permitido no puede representar con exactitud la fuente de información, en cuyo caso estamos obligados a realizar *compresión de datos* con pérdida.
- Transmisión de información a una velocidad mayor que la capacidad del canal.

Por tanto, la teoría de distorsión de velocidad puede considerarse como una extensión natural de los teoremas de codificación de Shannon.

■ FUNCIÓN DE DISTORSIÓN DE VELOCIDAD

Considere una fuente discreta sin memoria definida por un alfabeto M-ario $X:\{x_i | i = 1, 2, ..., M\}$, que consiste en un conjunto de símbolos estadísticamente independientes junto con las probabilidades de símbolo asociadas $\{p_i | i = 1, 2, ..., M\}$. Sea R la velocidad de código promedio en bits por palabra de código. Las palabras de código de la representación se toman de otro alfabeto $Y:\{y_j | j = 1, 2, ..., N\}$. El teorema de codificación de fuente establece que este segundo alfabeto proporciona una representación perfecta de la fuente siempre que $R > H$, donde H es la entropía de la fuente. Pero si estamos forzados a tener $R < H$, entonces existe una distorsión inevitable y, en consecuencia, pérdida de información.

Sea $p(x_i, y_j)$ la probabilidad conjunta de ocurrencia del símbolo de fuente x_i y símbolo de representación y_j. De acuerdo con la teoría de la probabilidad, tenemos

$$p(x_i, y_j) = p(y_j | x_i) p(x_i) \tag{9.127}$$

donde $p(y_j | x_i)$ es una probabilidad de transición. Sea $d(x_i, y_j)$ una medida del costo en el que se incurre al representar el símbolo de fuente x_i, mediante el símbolo y_j; la cantidad $d(x_i, y_j)$ se conoce como una *medida de distorsión de una letra*. El promedio estadístico de $d(x_i, y_j)$ sobre todos los posibles símbolos de fuente y símbolos de representación está dado por

$$\bar{d} = \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i) p(y_j | x_i) d(x_i, y_j) \tag{9.128}$$

Advierta que la distorsión promedio $\bar{d}$ es una función continua no negativa de las probabilidades de transición $p(y_j | x_i)$ que están determinadas por el par codificador-decodificador de fuente.

Se afirma que una asignación de probabilidad condicional $p(y_j | x_i)$ será *admisible* D si y sólo si la distorsión promedio $\bar{d}$ es menor o igual a algún valor aceptable D. El conjunto de todas las asignaciones de probabilidad condicional admisibles D se denota por medio de

$$P_D = \left\{ p(y_j | x_i) : \bar{d} \leq D \right\} \tag{9.129}$$

En cada conjunto de probabilidades de transición, tenemos una información mutua

$$I(X;Y) = \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i) p(y_j | x_i) \log \left(\frac{p(y_j | x_i)}{p(y_j)} \right) \tag{9.130}$$

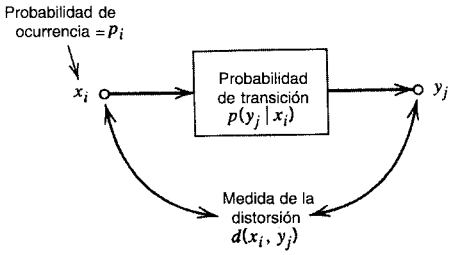


FIGURA 9.22 Resumen de la teoría de la distorsión de velocidad.

Una *función de distorsión de velocidad* $R(D)$ se define como *la velocidad de codificación más pequeña posible para la cual se garantiza que la distorsión promedio no superará* D. Sea P_D el conjunto al cual pertenece la probabilidad $p(y_j|x_i)$ para una D preestablecida. En ese caso, para una D fija escribimos[15]

$$R(D) = \min_{p(y_j|x_i)\in P_D} I(X;Y) \tag{9.131}$$

sujeta a la restricción

$$\sum_{j=1}^{N} p(y_j \mid x_i) = 1 \quad \text{para } i = 1, 2, \ldots, M \tag{9.132}$$

La función de distorsión de velocidad $R(D)$ se mide en unidades de bits si se usa el logaritmo de base 2 en la ecuación (9.130). Intuitivamente, esperamos que la distorsión D disminuya cuando aumente la función de distorsión de velocidad $R(D)$. Podemos afirmar inversamente que tolerar una gran distorsión D permite el uso de una velocidad más pequeña para la codificación y/o transmisión de la información.

La figura 9.22 resume los parámetros principales de la teoría de la distorsión de velocidad. En particular, dados los símbolos de fuente $\{x_i\}$ y sus probabilidades $\{p_i\}$ y dada la definición de la medida de distorsión de una sola letra $d(x_i, y_j)$, el cálculo de la función de distorsión de velocidad $R(D)$ incluye determinar la asignación de probabilidad condicional $p(y_j|x_i)$ sujeta a ciertas restricciones impuestas en $p(y_j|x_i)$. Éste es un problema variacional cuya solución desafortunadamente no es directa en general.

▶ EJEMPLO 9.14 Fuente gaussiana

Considere una fuente gaussiana sin memoria en tiempo discreto con media cero y varianza σ^2. Sea x el valor de la muestra generada por tal fuente y y una versión cuantizada de x que permite una representación finita de ella. La *distorsión de error cuadrático*

$$d(x,y) = (x-y)^2$$

proporciona una medida de la distorsión que se usa ampliamente para alfabetos continuos. La función de distorsión de velocidad para la fuente gaussiana con distorsión de error cuadrático, según se describe aquí, está dada por

$$R(D) = \begin{cases} \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right), & 0 \le D \le \sigma^2 \\ 0, & D > \sigma^2 \end{cases} \tag{9.133}$$

En este caso vemos que $R(D) \to \infty$ cuando $D \to 0$, y $R(D) = 0$ para $D = \sigma^2$. ◀

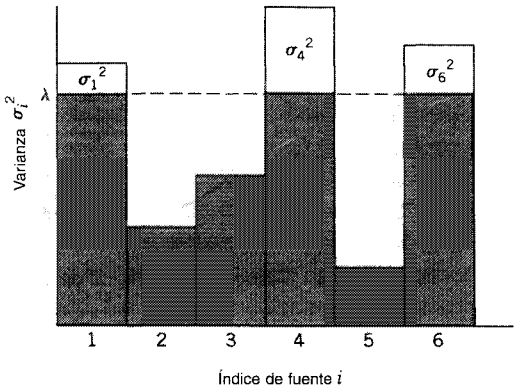


FIGURA 9.23 Imagen inversa del llenado de agua para un conjunto de procesos gaussianos en paralelo.

■ EJEMPLO 9.15 Conjunto de fuentes gaussianas en paralelo

Considere a continuación un conjunto de N variables aleatorias gaussianas independientes $\{X_i\}_{i=1}^N$, donde X_i tiene media cero y varianza σ_i^2. Utilizando la medida de distorsión

$$d = \sum_{i=1}^{N} (x_i - \hat{x}_i)^2$$

y con base en el resultado del ejemplo 9.14, podemos expresar la función de distorsión de velocidad para el conjunto de fuentes gaussianas en paralelo descrito aquí como

$$R(D) = \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\sigma_i^2}{D_i}\right) \tag{9.134}$$

donde D_i está a su vez definida por

$$D_i = \begin{cases} \lambda & \text{si } \lambda < \sigma_i \\ \sigma_i^2 & \text{si } \lambda \geq \sigma_i^2 \end{cases} \tag{9.135}$$

y la constante λ se elige de manera que satisfaga la condición

$$\sum_{i=1}^{N} D_i = D \tag{9.136}$$

Las ecuaciones (9.135) y (9.136) pueden interpretarse como un tipo de "llenado de agua a la inversa", como se ilustra en la figura 9.23. En primer lugar, elegimos una constante λ y únicamente el subconjunto de variables aleatorias cuyas varianzas exceden la constante λ. No se utilizan bits para describir el subconjunto restante de variables aleatorias cuyas varianzas son menores que la constante λ. ◀

9.14 *Compresión de datos*

La teoría de la distorsión de velocidad nos lleva de manera natural a considerar la idea de la *compresión de datos*, que implica una reducción a propósito e inevitable en el contenido de información de los datos provenientes de una fuente continua o discreta. Específicamente, es posible que consideremos a un *compresor de datos* o *compresor de señales* como un dispositivo que alimenta un código con el menor número de símbolos para la representación de la salida de la fuente, sujeta a una *distorsión* permisible o aceptable. El compresor de datos retiene de ese modo el contenido de información esencial de la salida de la fuente al borrar detalles finos de una manera deliberada pero controlada. Por tanto, la compresión de datos es una operación *con*

pérdidas a causa de que se reduce la entropía de la fuente (es decir, se pierde información), sin que importe el tipo de fuente que se está considerando.

En el caso de una fuente discreta, la razón para utilizar la compresión de datos radica en codificar la salida de la fuente a una velocidad menor que su entropía. Al hacerlo de ese modo, se viola el teorema de codificación de fuente, lo que significa que *ya no es posible* la reproducción exacta de los datos originales.

En el caso de una fuente continua, la entropía es infinita, y por ello debe utilizarse siempre un código de compresión de señales para codificar la salida de la fuente a una velocidad finita. En consecuencia, es imposible codificar digitalmente una señal analógica con un número finito de bits sin producir cierta distorsión. Este planteamiento concuerda perfectamente con la idea de modulación por codificación de pulsos, la cual se estudió en el capítulo 3. Ahí se demostró que la cuantización, la cual resulta básica para el proceso de conversión analógico-digital en la modulación por codificación de pulsos, siempre introduce distorsión (conocida como ruido de cuantización) en la señal transmitida. Por tanto, un cuantizador puede considerarse como un compresor de señales.

Los cuantizadores uniformes y no uniformes que se consideraron en el capítulo 3, se dice que son *escalares* porque trabajan con una señal analógica (es decir, salida de fuente continua) a la vez. Cada muestra se convierte en un valor cuantizado, siendo la conversión independiente de muestra a muestra. Un cuantizador escalar es un compresor de señales bastante simple, lo que lo hace atractivo en el uso práctico. Sin embargo, puede ofrecer un desempeño sorprendentemente bueno; esto es especialmente el caso si se recurre a la cuantización no uniforme.

Hay otra clase de estos elementos conocidos como *cuantizadores vectoriales* que utilizan bloques de muestras consecutivas de la salida de la fuente para formar vectores, cada uno de los cuales se trata como una sola entidad. La operación esencial en un cuantizador vectorial es la cuantización de un vector aleatorio[16] codificándolo como una palabra de código binaria. El vector se codifica al compararlo con un *registro de código* consistente en un conjunto de vectores de referencia almacenados conocidos como *vectores de código* o *patrones*. Cada patrón en el registro de código se emplea para representar vectores de entrada que mediante el codificador se identifican similares al patrón particular, sujeto a la maximización de un criterio de fidelidad apropiado. El proceso de codificación en un cuantizador vectorial puede considerarse consecuentemente como una *operación de vinculación de patrones*.

Sea N el número de vectores de código en el registro de código, k la dimensión de cada vector (es decir, el número de muestras en cada patrón) y r la velocidad de transmisión codificada en bits por muestra. Estos parámetros se relacionan del modo siguiente:

$$r = \frac{\log_2 N}{k} \tag{9.137}$$

En ese caso, suponiendo que el tamaño del registro de código es suficientemente grande, la relación señal a ruido (SNR) de cuantización para el cuantizador vectorial está dada por

$$10 \log_{10}(\text{SNR}) = 6\left(\frac{\log_2 N}{k}\right) + C_k \ dB \tag{9.138}$$

donde C_k es una constante (expresada en dB) que depende de las dimensiones de k. De acuerdo con la ecuación (9.138), la SNR para un cuantizador vectorial aumenta considerablemente a la tasa de $6/k$ dB por cada duplicación del tamaño del registro de código. De manera equivalente, podemos señalar que la SNR se incrementa en 6 dB por aumento unitario en la tasa (bits por muestra) como en la PCM estándar utilizando un cuantizador escalar uniforme. La ventaja del cuantizador vectorial sobre el escalar es que su término constante C_k tiene un valor superior, pues el cuantizador vectorial explota de manera óptima las correlaciones entre las muestras que constituyen un vector. De modo específico, la constante $xxxxC_k$ aumenta con la dimensión k, tendiendo a un límite último de distorsión de velocidad para una fuente de información determinada. Sin embargo, la mejora en la SNR se logra a costa de aumentar la complejidad de la

codificación, la cual crece exponencialmente con la dimensión k para una tasa especificada r. Desafortunadamente, éste es el principal obstáculo para el uso amplio en la práctica de la cuantización vectorial. A pesar de eso, en ciertas aplicaciones, el aspecto de la complejidad computacional se mitiga al explotar la capacidad de la tecnología VLSI para concentrar un procesador de señales altamente complejo en un chip de silicio. Por ejemplo, esto es precisamente lo que se hace en el uso del modelado predictivo lineal por excitación de código (CELP) del habla en los sistemas de comunicación inalámbricos del tipo CDMA, a saber, el sistema IS-95. A partir de la descripción que se presentó del CELP en la sección 8.9, es claro que el modelado del habla de este último elemento es un ejemplo de cuantización vectorial.

9.15 *Resumen y discusión*

En este capítulo establecimos cuatro límites fundamentales sobre los diferentes aspectos de un sistema de comunicación. Los límites se incorporan en el teorema de codificación de fuente, el teorema de codificación de canal, el teorema de capacidad de información y la función de distorsión de velocidad.

El *teorema de codificación de fuente*, primer teorema de Shannon, proporciona la herramienta matemática para valorar la *compactación de datos*, es decir, la *compresión sin pérdida* de datos generados por una fuente discreta sin memoria. El teorema nos indica que podemos hacer el número promedio de elementos de código binarios (bits) por símbolo de fuente tan pequeño, aunque no más que, la entropía de la fuente medida en bits. La *entropía* de una fuente es una función de las probabilidades de los símbolos de la fuente que constituyen el alfabeto de esta misma. Puesto que la entropía es una medida de incertidumbre, resulta máxima cuando la distribución de probabilidad asociada genera incertidumbre máxima.

El *teorema de codificación de canal*, segundo teorema de Shannon, es el resultado tanto más sorprendente como más importante de la teoría de la información. En un *canal simétrico binario*, el teorema de codificación de canal nos indica que para toda *tasa de código* r menor que o igual a la *capacidad del canal* C existen códigos tales que la probabilidad promedio de error es tan pequeña como queramos. Un canal simétrico binario es la forma más simple de un canal discreto sin memoria. Es simétrico porque la probabilidad de recibir un 1 si se envió un 0 es la misma que la probabilidad de recibir un 0 si se envió un 1. La probabilidad de que ocurra un error se denomina *probabilidad de transición*. La probabilidad de transición p se determina no sólo mediante el ruido aditivo a la salida del canal, sino también a partir del tipo de receptor que se utiliza. El valor de p define en forma única la capacidad C del canal.

El tercer teorema notable de Shannon, el *teorema de la capacidad de información*, nos dice que hay un máximo para la velocidad a la cual todo sistema de comunicación puede operar confiablemente (es decir, sin errores) cuando el sistema sufre una restricción de potencia. Esta velocidad máxima se conoce como *capacidad de información*, medida en bits por segundo. Cuando el sistema opera a una tasa mayor que la capacidad de información, está condenado a una elevada probabilidad de error, independientemente de la elección del conjunto de señales utilizado para la transmisión o del receptor que se emplea para procesar la señal recibida.

Por último, la función de distorsión de velocidad proporciona la herramienta matemática para la compresión de señales (es decir, resolver el problema de la codificación de fuente con un criterio de fidelidad): la función de distorsión de la tasa puede aplicarse a una fuente discreta sin memoria o continua.

Cuando la salida de una fuente de información se comprime de una manera en la que no existen pérdidas, la secuencia de datos que resulta usualmente contiene bits redundantes. Estos últimos pueden eliminarse utilizando un algoritmo sin pérdidas, como la codificación de Huffman o el algoritmo Lempel-Ziv para compactación de datos. Por ello es posible hablar de la compresión de datos seguida por la compactación de datos como dos constituyentes de la *disección de la*

codificación de fuente, la cual se denomina de esa manera debido a que se refiere exclusivamente a las fuentes de información. En algunas aplicaciones de codificación de fuente, tenemos un tercer constituyente, a saber, el *encriptado de datos,* el cual sigue a la compactación de datos. El propósito del encriptado es disfrazar la secuencia de datos (bits) de tal manera que éstos no tengan significado para un receptor desautorizado. Algunos aspectos básicos de la *criptografía,* los cuales abarcan tanto el encriptado como el desencriptado, surgen de una manera bastante natural de la teoría de la información, como se explica en el apéndice 5. Otros aspectos que resultan de la criptografía se analizan también en ese apéndice.

Cabe un último comentario. La teoría de información de Shannon, según se presenta en este capítulo, ha estado por completo en el contexto de las fuentes y canales sin memoria. La teoría puede extenderse para abordar fuentes y canales con *memoria,* en cuyo caso un símbolo de interés depende de los símbolos precedentes; sin embargo, el nivel de exposición que se necesita para lo anterior se encuentra más allá del objetivo de este libro.[17]

NOTAS Y BIBLIOGRAFÍA

1. De acuerdo con Lucky [96], la primera mención que Shannon hizo del término *teoría de la información* ocurrió en una nota de 1945 titulada "Una teoría matemática de la criptografía". Es bastante curioso que el término nunca se usó en el artículo clásico de 1948 de Shannon, el cual puso los cimientos de la teoría de la información. En lo que se refiere a un tratamiento introductorio de esta teoría, vea el capítulo 2 de Lucky [96] y el artículo de Wyner [407]; consulte también los libros de Adámek [3], Hamming [60] y Abramson [2]. Para tratamientos más avanzados del tema, vea los libros de Cover y Thomas [33], Blahut [23], McEliece [103] y Gallager [46]. Con respecto a una colección de artículos sobre el desarrollo de la teoría de la información (incluso el artículo clásico de 1948 de Shannon) vea Slepian [141]. Para una colección de artículos publicados por Shannon, consulte Sloane y Wyner [142].

2. El uso de una medida logarítmica de información fue sugerido por primera vez por Hartley [278]; sin embargo, Hartley utilizó logaritmos de base 10.

3. En la física estadística, la entropía de un sistema físico se define por medio de (Reif, p.147)

$$\mathscr{S} = k \log \Omega$$

donde k es la constante de Boltzmann, Ω es el número de estados accesibles para el sistema y log denota el logaritmo natural. Esta entropía tiene las dimensiones de energía debido a que su definición implica la constante k. En particular, proporciona una *medida cuantitativa del grado de aleatoriedad del sistema.* Comparando la entropía de la física estadística con la teoría de la información, vemos que tienen una forma similar. Para una explicación detallada de la relación entre ellas, vea Pierce ([120], pp. 184-207) y Brillouin [27].

4. En cuanto a la prueba original del teorema de codificación de fuente vea Shannon [368]. Una prueba general del teorema de codificación de fuente se proporciona también en los siguientes libros: Viterbi y Omura ([162], pp. 13-19), McEliece ([103], capítulo 3) y Gallager ([46], pp. 38-55). El teorema de codificación de fuente se conoce también en la literatura como el *teorema de codificación sin ruido,* en el sentido de que establece la condición para que sea posible la codificación sin errores.

5. Para la prueba de desigualdad de Kraft-McMillan, vea Cover y Thomas ([33], pp. 82-84), Blahut ([24], pp. 298-299) y McEliece ([103], pp. 239-240). En lo que respecta a una prueba de la ecuación (9.23), revise Cover y Thomas ([33], pp. 87-88), Blahut ([24], pp. 300-301) y McEliece ([103], pp. 241-242).

6. El código de Huffman recibe ese nombre en honor a su inventor: D.A. Huffman [280]. Para una explicación legible del código de Huffman y su uso en la compactación de datos vea Adámek [3].

7. Los artículos originales sobre el algoritmo Lempel-Ziv, son Ziv y Lempel [411, 412]. En cuanto a descripciones fáciles de comprender del algoritmo de Lempel-Ziv consulte Lucky ([96], pp. 118-

122), Blahut ([24], pp. 314-319) y Gitlin, Hayes y Weinstein ([52], pp. 120-122). En lo relativo a una aplicación del algoritmo Lempel-Ziv a la compactación de texto en inglés, vea Lucky ([96], pp. 122-128) y el artículo de Welch [396]; vea también el artículo general de Weiss y Shremp [395].

8. El teorema de codificación del canal se conoce también como *teorema de codificación de ruido*. La prueba original del teorema se proporciona en Shannon [369]. Una prueba del teorema se ofrece también en Hamming ([60], capítulos 9 y 10) con suficiente detalle, de manera que conforma una apreciación general de los resultados relevantes. La segunda parte del teorema se conoce en la literatura como el *inverso del teorema de codificación*. Una prueba de este último se presenta en las siguientes referencias: Viterbi y Omura ([162], pp. 28-34) y Gallager ([46], pp. 76-82).

9. La cantidad

$$\int_{-\infty}^{\infty} f_Y(x) \log_2 \left(\frac{f_X(x)}{f_Y(x)} \right) dx$$

en el primer miembro de la ecuación (9.70) recibe el nombre de *entropía relativa* o *divergencia de Kullback-Leibler* entre las funciones de densidad de probabilidad $f_X(x)$ y $f_Y(x)$; consulte Kullback [81].

10. El teorema de Shannon de la capacidad de la información se conoce también en la literatura como la *ley Shannon-Hartley* en reconocimiento al trabajo previo de Hartley acerca de la transmisión de la información (Hartley, [278]). En particular, este último autor demostró que la cantidad de información que puede transmitirse por un canal determinado es proporcional al producto del ancho de banda del canal y el tiempo de operación.

11. Una clara exposición del empaquetamiento de esferas se presenta en Cover y Thomas ([33], pp. 242-245); vea también Wozencraft y Jacobs ([172], pp. 323-341).

12. Las partes *a* y *b* de la figura 9.18 siguen de las partes correspondientes de la figura 6.2 en el libro de Frey.

13. Para un tratamiento riguroso de la capacidad de información de un canal ruidoso de color, vea Gallager [46]. La idea de sustituir el modelo de canal de la figura 9.19*a* con el de la figura 9.19*b* se explica en Gitlin, Hayes y Weinstein [52].

14. Con respecto a un tratamiento completo de la teoría de la distorsión de la tasa, vea el libro de Berger [15]; este tema se aborda también con un poco de menos detalle en Cover y Thomas [33], McEliece [103] y Gallager [46].

15. Para la deducción de la ecuación (9.131) vea Cover y Thomas ([33], p. 345). Un algoritmo para el cálculo de la función de la distorsión de velocidad $R(D)$ definida en la ecuación (9.131) se describe en Blahut ([23], pp. 220-221) y Cover y Thomas ([33], pp. 364-367).

16. En cuanto a artículos anteriores sobre la cuantización vectorial, consulte a Gersho y Linde, y vea a Buzo y Gray [307]. En lo referente a un repaso instructivo de la cuantización de vectores, vea a Gray [268]. La ecuación (9.138), que define la SNR para un cuantizador vectorial, la explican Gersho y Cuperman. En lo relativo a un tratamiento completo de la cuantización de vectores, consulte el libre de Gersho y Gray [50].

17. Para una explicación detallada de los canales discretos con memoria, vea Gallager ([46], pp. 97-112) y Ash ([8], pp. 211-229).

PROBLEMAS

Entropía

9.1 Sea p la probabilidad de cierto evento. Grafique la cantidad de información ganada por la ocurrencia de este evento para $0 \leq p \leq 1$.

9.2 Una fuente emite uno de cuatro posibles símbolos durante cada intervalo de transmisión de señales. Los símbolos ocurren con las probabilidades:

$$p_0 = 0.4$$
$$p_1 = 0.3$$
$$p_2 = 0.2$$
$$p_3 = 0.1$$

Determine la cantidad de información ganada al observar la fuente que emite cada uno de estos símbolos.

9.3 Una fuente emite uno de cuatro símbolos s_0, s_1, s_2 y s_3 con probabilidades 1/3, 1/6, 1/4 y 1/4, respectivamente. Los símbolos sucesivos emitidos por la fuente son estadísticamente independientes. Calcule la entropía de la fuente.

9.4 Considere que X representa el resultado de un solo lanzamiento de un dado sin alterar. ¿Cuál es la entropía de X?

9.5 La función de muestreo de un proceso gaussiano de media cero y varianza unitaria se muestrea de manera uniforme y luego se aplica a un cuantizador uniforme que tiene la característica de amplitud de entrada-salida que se muestra en la figura P9.5. Calcule la entropía de la salida del cuantizador.

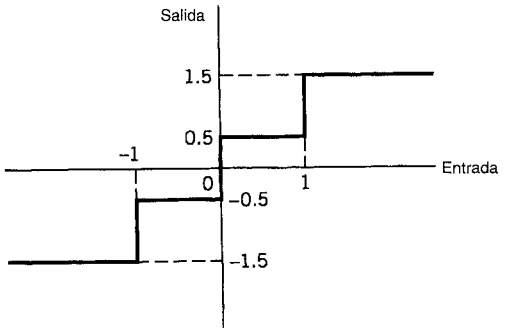


FIGURA P9.5

9.6 Considere una fuente discreta sin memoria con alfabeto de fuente $\mathscr{S} = \{s_0, s_1, \ldots, s_{K-1}\}$ y estadísticas de fuente $\{p_0, p_1, \ldots, p_{K-1}\}$. La n-ésima extensión de esta fuente es otra fuente discreta sin memoria con alfabeto de fuente $\mathscr{S}^n = \{\sigma_0, \sigma_1, \ldots, \sigma_{M-1}\}$, donde $M = K^n$. Considere que $P(\sigma_i)$ denota la probabilidad de σ_i.

a) Demuestre que

$$\sum_{i=0}^{M-1} P(\sigma_i) = 1$$

que es lo que se esperaba.

b) Demuestre que

$$\sum_{i=0}^{M-1} P(\sigma_i) \log_2\left(\frac{1}{p_{i_k}}\right) = H(\mathscr{S}), \qquad k = 1, 2, \ldots, n$$

donde p_{i_k} es la probabilidad del símbolo s_{i_k}, y $H(\mathscr{S})$ es la entropía de la fuente original.

c) Por tanto, demuestre que

$$H(\mathscr{S}^n) = \sum_{i=0}^{M-1} P(\sigma_i) \log_2 \frac{1}{P(\sigma_i)}$$
$$= nH(\mathscr{S})$$

9.7 Considere una fuente discreta sin memoria con alfabeto de fuente $\mathscr{S} = \{s_0, s_1, s_2\}$ y estadísticas de fuente {0.7, 0.15, 0.15}.

a) Calcule la entropía de la fuente.

b) Calcule la entropía de la extensión de segundo orden de la fuente.

9.8 Posiblemente sea una sorpresa, pero el número de bits necesario para almacenar texto es mucho menor que el requerido para almacenar su equivalente hablado. ¿Puede usted explicar la razón de ello?

Compactación de datos

9.9 Considere una fuente discreta sin memoria cuyo alfabeto consiste en K símbolos igualmente probables.

a) Explique por qué el uso de un código de longitud fija para la representación de una fuente de este tipo es aproximadamente tan eficiente como puede serlo cualquier código.

b) ¿Qué condiciones tiene que satisfacer K y la longitud de palabra de código para que la eficiencia de la codificación sea de 100%?

9.10 Considere los cuatro códigos listados a continuación:

Símbolo	*Código I*	*Código II*	*Código III*	*Código IV*
s_0	0	0	0	00
s_1	10	01	01	01
s_2	110	001	011	10
s_3	1110	0010	110	110
s_4	1111	0011	111	111

a) Dos de estos cuatro son códigos de prefijo. Identifíquelos y construya sus árboles de decisión individuales.

b) Aplique la desigualdad de Kraft-McMillan a los códigos I, II, III y IV. Discuta sus resultados a la luz de los que se obtuvieron en la parte *a*.

9.11 Considere una secuencia de letras del alfabeto inglés con sus probabilidades de ocurrencia como en las dadas a continuación:

Letra	*a*	*i*	*l*	*m*	*m*	*o*	*p*	*y*
Probabilidad	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1

Calcule los diferentes códigos de Huffman para este alfabeto. En un caso, mueva un símbolo combinado en el procedimiento de codificación tan alto como sea posible, y en el segundo caso, muévalo lo más bajo posible. Por tanto, para cada uno de los dos códigos determine la longitud promedio de palabra de código y la varianza de esta misma longitud con respecto al conjunto de letras.

9.12 Una fuente discreta sin memoria tiene un alfabeto de siete símbolos cuyas probabilidades de ocurrencia se describen a continuación:

Símbolo	s_0	s_1	s_2	s_3	s_4	s_5	s_6
Probabilidad	0.25	0.25	0.125	0.125	0.125	0.0625	0.0625

Calcule el código de Huffman para esta fuente, moviendo un "símbolo combinado" tan alto como sea posible. Explique por qué el código fuente calculado tiene una eficiencia del 100%.

9.13 Considere una fuente discreta sin memoria con alfabeto $\{s_0, s_1, s_2\}$ y estadísticas {0.7, 0.15, 0.15} para su salida.

a) Aplique el algoritmo de Huffman para esta fuente. Por lo tanto, demuestre que la longitud promedio del código de Huffman es igual a 1.3 bits/símbolo.

b) Deje que la fuente se extienda hasta el orden dos. Aplique el algoritmo de Huffman a la fuente extendida resultante y demuestre que la longitud promedio del nuevo código es igual a 1.1975 bits/símbolo.

c) Compare la longitud promedio de palabra de código calculada en la parte *b* con la entropía de la fuente original.

9.14 La figura P9.14 muestra un árbol de Huffman. ¿Cuál es la palabra de código para cada uno de los símbolos A, B, C, D, E, F y G representada por este árbol de Huffman? ¿Cuáles son sus longitudes de palabra de código individuales?

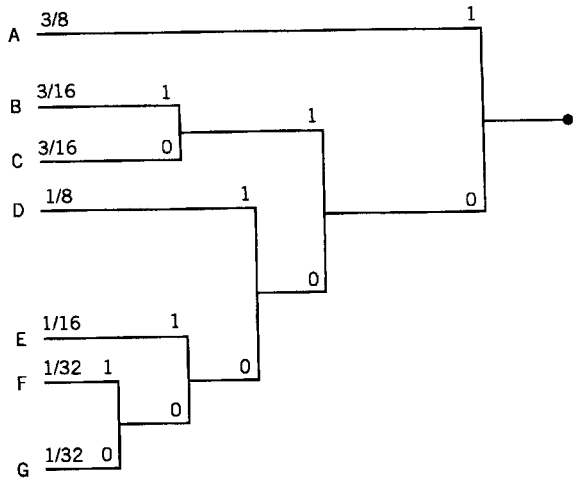


FIGURA P9.14

9.15 Una computadora ejecuta cuatro instrucciones que se designan mediante las palabras de código (00, 01, 10, 11). Suponiendo que las instrucciones se usan independientemente con las probabilidades (1/2, 1/8, 1/8, 1/4), calcule el porcentaje mediante el cual el número de bits utilizados para las instrucciones puede reducirse mediante el uso de un código de fuente óptimo. Construya un código de Huffman para realizar la reducción.

9.16 Considere la siguiente secuencia binaria

$$11101001100010110100\ldots$$

Utilice el algoritmo de Lempel-Ziv para codificar esta secuencia. Suponga que los símbolos binarios 0 y 1 se encuentran ya en el registro de código.

Canal simétrico binario

9.17 Considere el diagrama de probabilidad de transición del canal simétrico binario que se muestra en la figura 9.8. Los símbolos binarios de entrada 0 y 1 ocurren con igual probabilidad. Determine las probabilidades de los símbolos binarios 0 y 1 que aparecen a la salida del canal.

9.18 Repita el cálculo del problema 9.17 suponiendo que los símbolos de entrada 0 y 1 ocurren con probabilidades 1/4 y 3/4, respectivamente.

Información mutua y capacidad del canal

9.19 Considere un canal simétrico binario caracterizado por la probabilidad de transición p. Grafique la información mutua del canal como una función de p_1, la probabilidad *a priori* del símbolo 1 a la entrada del canal; realice sus cálculos para la probabilidad de transición $p = 0, 0.1, 0.2, 0.3, 0.5$.

9.20 La figura 9.10 describe la variación de la capacidad de canal de un canal simétrico binario con la probabilidad de transición p. Recurra a los resultados del problema 9.19 para explicar esta variación.

9.21 Considere el canal simétrico binario descrito en la figura 9.8. Sea p_0 la probabilidad de enviar el símbolo binario $x_0 = 0$, y sea $p_1 = 1 - p_0$ la probabilidad de enviar el símbolo binario $x_1 = 1$. Considere que p denota la probabilidad de transición del canal.

a) Demuestre que la información mutua entre la entrada y la salida del canal está dada por

$$I(\mathcal{X};\mathcal{Y}) = \mathcal{H}(z) - \mathcal{H}(p)$$

donde

$$H(z) = z\log_2\left(\frac{1}{z}\right) + (1-z)\log_2\left(\frac{1}{1-z}\right)$$
$$z = p_0 p + (1-p_0)(1-p)$$

y

$$H(p) = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right)$$

b) Demuestre que el valor de p_0 que hace máxima a $I(\mathcal{X}, \mathcal{Y})$ es igual a 1/2.

c) Por tanto, compruebe que la capacidad del canal es igual a

$$C = 1 - H(p)$$

9.22 Dos canales simétricos binarios se conectan en cascada, en la forma que se indica en la figura P9.22. Determine la capacidad total del canal de la conexión en cascada, suponiendo que ambos canales tienen el mismo diagrama de probabilidad de transición que se muestra en la figura 9.8.

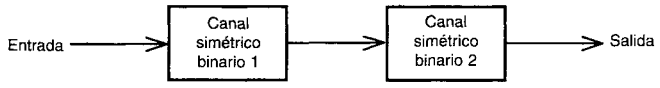


FIGURA P9.22

9.23 El *canal de borradura binaria* tiene dos entradas y tres salidas, como se describe en la figura P9.23. Las entradas se marcan con 0 y 1, y las salidas con 0, 1 y e. Una fracción a de los bits entrantes se borra por medio del canal. Determine la capacidad de este mismo.

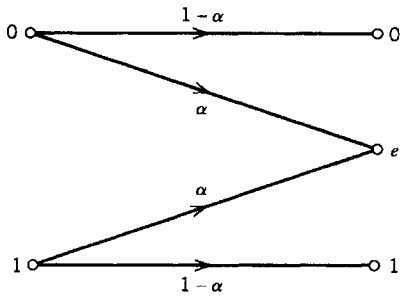


FIGURA P9.23

9.24 Considere un sistema de comunicación digital que utiliza un *código de repetición* para la codificación/decodificación del canal. En particular, cada transmisión se repite n veces, donde $n = 2m + 1$ es un entero impar. El decodificador opera del modo siguiente. Si en un bloque de n bits recibido, el número de ceros supera al número de unos, el decodificador decide a favor de un 0. En otro caso, lo hace a favor de un 1. Ocurre un error cuando $m + 1$ o más transmisiones fuera de $n = 2m + 1$ son incorrectas. Suponga un canal simétrico binario.

a) Para $n = 3$, demuestre que la probabilidad promedio de error está dada por

$$P_e = 3p^2(1-p) + p^3$$

donde p es la probabilidad de transición del canal.

b) Para $n = 5$, demuestre que la probabilidad promedio de error está dada por

$$P_e = 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

c) Por tanto, para el caso general, deduzca que la probabilidad promedio de error se determina mediante

$$P_e = \sum_{i=m+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Entropía diferencial

9.25 Sean $X_1, X_2, ..., X_n$ los elementos de un vector gaussiano $\mathbf{X}$. Las X_i son independientes con media μ_i y varianza σ^2_i, $i = 1, 2, ..., n$. Demuestre que la entropía diferencial del vector $\mathbf{X}$ es igual a

$$h(\mathbf{X}) = \frac{n}{2} \log_2 \left[2\pi e (\sigma_1^2 \sigma_2^2 \ldots \sigma_n^2)^{1/n} \right]$$

¿A qué se reduce $h(\mathbf{X})$ si las varianzas son iguales?

9.26 Una variable aleatoria continua X está restringida a una magnitud pico M; es decir, $-M < X < M$.

a) Compruebe que la entropía diferencial de X es un máximo cuando ésta se distribuye uniformemente, según se indica mediante

$$f_X(x) = \begin{cases} 1/2M, & -M < x \leq M \\ 0, & \text{en otro caso} \end{cases}$$

b) Demuestre que la entropía diferencial máxima de X es $\log_2 2M$.

9.27 Pruebe las propiedades dadas en las ecuaciones de la (9.79) a la (9.81) para la información mutua $I(X; Y)$.

9.28 Considere la variable aleatoria continua Y definida por

$$Y = X + N$$

donde X y N son independientes estadísticamente. Demuestre que la entropía diferencial condicional de Y, dada X, es igual a

$$h(Y|X) = h(N)$$

donde $h(N)$ es la entropía diferencial de N.

Capacidad de información

9.29 Un canal de calidad de voz de la red telefónica tiene un ancho de banda de 3.4 kHz.

a) Calcule la capacidad de información del canal telefónico para una relación señal a ruido de 30 dB.

b) Calcule la relación señal a ruido mínima que se requiere para soportar la transmisión de información a través del canal telefónico a una tasa de 9,600 b/s.

9.30 Se alimentan datos alfanuméricos a una computadora desde una terminal remota a través de un canal telefónico de calidad de voz. El canal tiene un ancho de banda de 3.4 kHz y una relación de señal de salida a ruido de 20 dB. La terminal tiene un total de 128 símbolos. Suponga que éstos son igualmente probables y que las transmisiones sucesivas son estadísticamente independientes.

a) Calcule la capacidad de información del canal.

b) Calcule la velocidad de símbolos máxima para la cual es posible la transmisión sin errores por el canal.

9.31 Una imagen de televisión en blanco y negro puede considerarse como si estuviera compuesta de aproximadamente 3×10^5 elementos, cada uno de los cuales quizás ocupe uno de diez distintos niveles de brillantez con igual probabilidad. Suponga que: 1) la velocidad de transmisiones es de 30 cuadros de imagen por segundo, y 2) que la relación de señal a ruido corresponde a 30 dB.

Empleando el teorema de la capacidad de información, calcule el ancho de banda mínimo que se requiere para soportar la transmisión de la señal de video resultante.

Nota: como una cuestión de interés, las transmisiones de televisión comercial emplean en realidad un ancho de banda de 4.2 MHz, el cual encaja dentro del ancho de banda asignado de 6 MHz.

9.32 En este problema continuamos con el ejemplo 9.9. Suponga que la constelación estrechamente empaquetada de la figura 9.15*b* se escala hacia arriba de manera que la energía de la señal transmitida por símbolo se mantiene en el mismo valor promedio que la consumida por la constelación cuadrada QAM 64 de la figura 9.15*a*. Construya la nueva constelación que se produce a partir de este escalamiento. ¿Cómo se compara la tasa de error de bits de esta nueva constelación con la de la figura 9.15*a*? Justifique su respuesta.

9.33 La respuesta en magnitud al cuadrado de un canal de par trenzado puede modelarse como

$$|H(f)|^2 = \exp(-\alpha\sqrt{f})$$

La constante a está definida por

$$\alpha = \frac{kl}{l_0}$$

donde k es una constante que depende del calibre del alambre, l_0 es una longitud de la línea de referencia y l es la longitud real del par trenzado bajo estudio. La respuesta en magnitud al cuadrado del acoplamiento responsable de la NEXT tiene la forma

$$|H_{\text{NEXT}}(f)|^2 = \beta f^{3/2}$$

donde β es una constante que depende del tipo de cable utilizado.

Formule la expresión para la capacidad de información del canal dominado por la NEXT descrito aquí.

Compresión de datos

9.34 La ecuación (9.138) para la relación señal a ruido (SNR) de un cuantizador vectorial incluye la fórmula SNR de la ecuación (3.33) para la modulación por codificación de pulsos estándar como un caso especial para el cual $k = 1$. Justifique la validez de esta inclusión.

9.35 Todos los esquemas prácticos de compresión y transmisión de datos se ubican entre dos límites impuestos por la función de distorsión de velocidad y el teorema de capacidad del canal. Ambos teoremas incluyen la noción de información mutua, aunque de manera diferente. Comente acerca de los aspectos que surgen a partir de estos dos enunciados.

Experimento de computadora

9.36 En este problema volvemos al ejemplo 9.12, el cual tiene que ver con la transmisión de señales antípodas binarias codificadas por un canal con ruido blanco gaussiano aditivo (AWGN). Empezando con la ecuación (9.112) y con la teoría subyacente, desarrolle un paquete de software para calcular el mínimo E_b/N_0 requerido para una tasa de error de bits determinada, donde E_b es la energía de

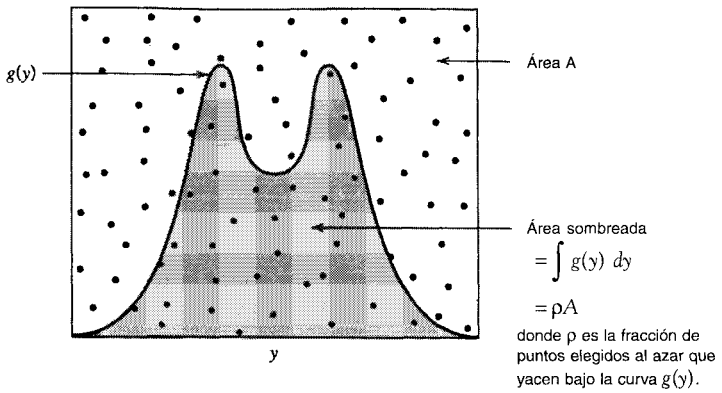


FIGURA P9.36

la señal por bit y $N_0/2$ es la densidad espectral del ruido. Por lo tanto, calcule los resultados que se grafican en las partes *a* y *b* de la figura 9.18.

Como se menciona en el ejemplo 9.12, el cálculo de la información mutua entre la entrada y la salida del canal se aproxima adecuadamente utilizando integración de Monte Carlo. Para explicar cómo funciona este método, considere una función g(y) que es difícil de muestrear aleatoriamente, lo cual es en realidad el caso en el problema presente. (En nuestro problema, la función g(y) representa el integrando complicado en la fórmula correspondiente a la entropía diferencial de la salida del canal.) Para el cálculo, proceda de la siguiente manera:

- Encuentre el área A que incluye la región de interés y que se muestrea con facilidad.
- Elija *N* puntos uniformemente al azar dentro del área A.

En ese caso el *teorema de integración de Monte Carlo* establece que la función integral g(y) con respecto a y es aproximadamente igual al área A multiplicada por la fracción de puntos que residen debajo de la curva g, como se ilustra en la figura P9.36. La exactitud de la aproximación mejora al crecer *N*.

CAPÍTULO 10

CODIFICACIÓN DE CONTROL DE ERRORES

Este capítulo es la secuencia natural del anterior sobre la teoría de la información de Shannon. En particular, en él se presentan técnicas de codificación de control de errores que ofrecen maneras diferentes de poner en práctica el teorema de codificación del canal de Shannon. Cada técnica de control de errores implica el uso de un codificador de canal en el transmisor y de un algoritmo de decodificación en el receptor.

Las técnicas de codificación de control de errores que se describen aquí incluyen las siguientes clases de códigos importantes:

- *Códigos de bloque lineales.*
- *Códigos cíclicos.*
- *Códigos convolucionales.*
- *Códigos compuestos ejemplificados por códigos turbo y códigos de verificación de paridad de baja densidad, y sus variantes irregulares.*

10.1 *Introducción*

La tarea que encara el diseñador de un sistema digital de comunicaciones es proveer una instalación al costo más conveniente para transmitir información desde un extremo del sistema a una velocidad y nivel de confiabilidad y calidad aceptables para el usuario en el otro extremo. Los dos parámetros principales del sistema con que cuenta el diseñador son la potencia de la señal transmitida y el ancho de banda del canal. Estos dos parámetros, junto con la densidad espectral de potencia del ruido del receptor, establecen la razón entre la energía de la señal por bit y la densidad espectral de potencia del ruido E_b/N_0. En el capítulo 6 mostramos que esta proporción determina singularmente la tasa de error de bits en un esquema de modulación particular. Las consideraciones prácticas suelen poner un límite en el valor que podemos asignar a E_b/N_0. De acuerdo con esto, en la práctica llegamos con frecuencia a un esquema de modulación y descubrimos que no es posible proporcionar una calidad de datos aceptable (es decir, un desempeño de errores suficientemente bajo). Para un E_b/N_0 fijo, la única opción práctica disponible para cambiar la calidad de los datos de problemática a aceptable consiste en utilizar la *codificación de control de errores*.

Otra motivación práctica para el empleo de la codificación radica en reducir la proporción E_b/N_0 requerida para una tasa de error de bits fija. Esta reducción en E_b/N_0 puede, a su vez, explotarse para reducir la potencia transmitida necesaria o reducir los costos de hardware requiriendo un tamaño de antena más pequeño en el caso de comunicaciones de radio.

El *control de errores*[1] para la integridad de datos puede efectuarse por medio de la *corrección de errores directa* (FEC). La figura 10.1*a* muestra el modelo de un sistema de comunicación digital que utiliza un procedimiento de este tipo. La fuente discreta genera información en la forma de símbolos binarios. El *codificador de canal* en el transmisor acepta los bits del mensaje y agrega redundancia de acuerdo con una regla preestablecida, produciendo de ese modo datos

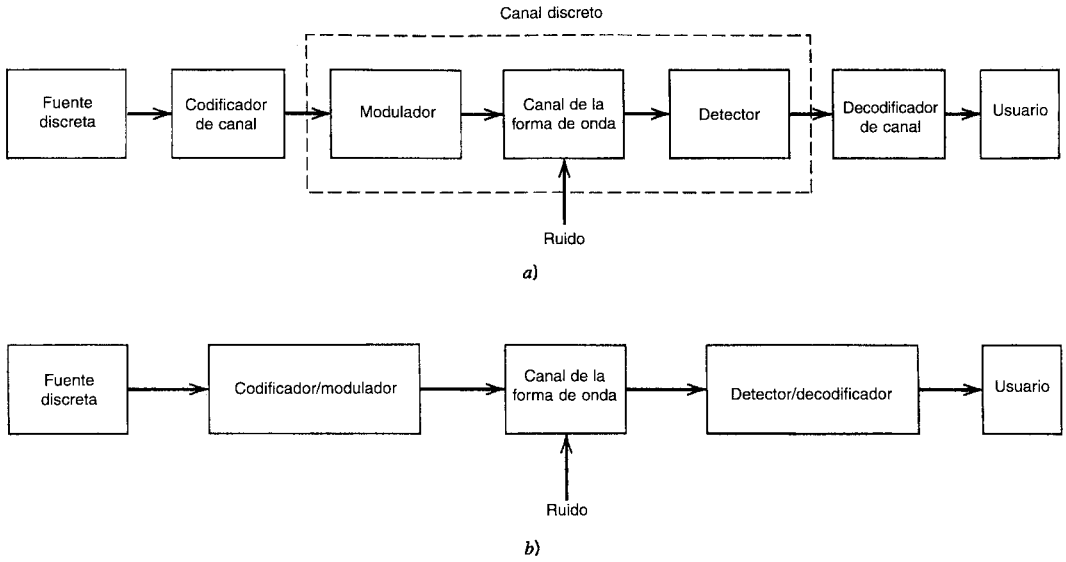


FIGURA 10.1 Modelos simplificados del sistema de comunicación. *a*) Codificación y modulación efectuadas por separado. *b*) Codificación y modulación combinadas.

codificados a una tasa de bits más alta. El *decodificador del canal* en el receptor explota la redundancia para decidir cuáles bits del mensaje se transmitieron realmente. La meta combinada del codificador y el decodificador de canal reside en minimizar el efecto del ruido del canal. Es decir, se minimiza el número de errores entre la entrada del codificador del canal (obtenida de la fuente) y la salida del decodificador del canal (entregada al usuario).

En un esquema de modulación fijo, la adición de redundancia en los mensajes codificados implica la necesidad de aumentar el ancho de banda de la transmisión. Además, el uso de la codificación de control de errores añade *complejidad* al sistema, en especial en la puesta en práctica de las operaciones de decodificación en el receptor. De ese modo, los compromisos de diseño en el uso de la codificación del control de errores para lograr un desempeño de errores aceptable incluye consideraciones del ancho de banda y la complejidad del sistema.

Existen muchos códigos diferentes de corrección de errores (con raíces en diversas disciplinas matemáticas) que podemos usar. Históricamente, estos códigos se han clasificado en *códigos de bloque* y *códigos de convolución*. El rasgo distintivo para esta clasificación particular es la presencia o ausencia de *memoria* en los codificadores de los dos códigos.

Para generar un código de bloque (n, k), el codificador del canal acepta información en bloques sucesivos de k bits; en cada bloque agrega $n - k$ bits redundantes que se relacionan algebraicamente con los k bits del mensaje, produciendo por ello un bloque codificado completo de n bits, donde $n > k$. El bloque de n bits se denomina una *palabra de código*, y n se llama la *longitud del bloque* del código. El codificador del canal produce bits a razón de $R_0 = (n/k)R_s$, donde R_s es la tasa de bits de la fuente de información. La razón adimensional $r = k/n$ se conoce como la *tasa de código*, donde $0 < r < 1$. La tasa de bits R_0, proveniente del codificador, se denomina *tasa de datos del canal*. Así, la tasa de código es una razón adimensional, en tanto que la tasa de datos producida por la fuente y del canal se miden en ambos casos en bits por segundo.

En un código de convolución, la operación de codificado puede verse como la *convolución en tiempo discreto* de la secuencia de entrada con la respuesta al impulso del codificador. La duración de la respuesta al impulso es igual a la memoria del codificador. En consecuencia, el codificador en un código convolucional opera sobre la secuencia del mensaje entrante, empleando una "ventana de deslizamiento" igual en duración a su propia memoria, lo cual, a su

vez, significa que en un código convolucional, a diferencia de un código de bloque, el codificador de canal acepta bits de mensaje como una secuencia continua, por lo que genera una secuencia de bits codificados a una tasa más alta.

En el modelo que se describe en la figura 10.1*a*, las operaciones de codificación de canal y modulación se efectúan por separado en el transmisor; lo mismo ocurre en las operaciones de detección y decodificación en el receptor. Sin embargo, cuando la eficiencia del ancho de banda es de interés principal, el método más efectivo para poner en práctica la corrección de codificación directa de control de error consiste en combinarla con la modulación como una sola función, según se muestra en la figura 10.1*b*. En un método de estas características, la codificación se redefine como un proceso de imponer ciertos patrones sobre la señal transmitida.

■ SOLICITUD DE REPETICIÓN AUTOMÁTICA

La corrección de errores de alimentación directa (FEC) se fundamenta en el uso controlado de redundancia en la palabra de código transmitida tanto para la *detección como corrección* de los errores en los que se incurre durante el curso de la transmisión por un canal ruidoso. Sin que importe si la decodificación de la palabra de código recibida es exitosa, no se realiza procesamiento adicional en el receptor. En consecuencia, las técnicas de codificación de canal adecuadas para la FEC sólo requieren un *enlace de una vía* entre el transmisor y el receptor.

Existe otro método conocido como *solicitud de repetición automática* (ARQ)[2] para resolver el problema del control de errores. La filosofía que subyace en la ARQ es bastante diferente de la correspondiente a la FEC. De modo específico, la ARQ utiliza la redundancia absolutamente con propósitos de *detección de errores*. En la detección de un error en una palabra de código transmitida, el receptor solicita una transmisión repetida de la palabra de código corrompida, lo cual necesita el uso de una *trayectoria de retorno* (es decir, un canal de realimentación). Como tal, la ARQ puede utilizarse únicamente en enlaces *semidúplex* o *dúplex completo*. En el enlace semidúplex, la transmisión de datos por el enlace puede realizarse en cualquier dirección, pero *no* en forma simultánea. Por otra parte, en el enlace dúplex completo, resulta posible que la transmisión de datos se lleve a cabo en forma simultánea por el enlace en ambas direcciones.

Un enlace semidúplex utiliza el esquema ARQ simple conocido como *estrategia de paro y espera*. En este método, un bloque de bits de mensaje se codifica en una palabra de código y se transmite por el canal. El transmisor se detiene después y espera para la realimentación desde el receptor. La señal de realimentación posiblemente corresponda al reconocimiento de la recepción correcta de la palabra de código o a una solicitud de transmisión de la palabra de código debido a un error en su decodificación. En último caso, el transmisor reenvía la palabra del código en cuestión antes de moverse hacia el siguiente bloque de bits de mensaje.

El problema de inactividad en la ARQ en paro y espera origina un reducido rendimiento de datos, el cual se alivia en otro tipo de ARQ conocida como *ARQ continua con retirada*. Esta segunda estrategia utiliza un enlace dúplex completo, lo cual permite al receptor enviar una señal de realimentación en tanto que el transmisor está comprometido en el envío de las palabras de código por el canal directo. De manera específica, el transmisor continúa enviando una sucesión de palabras de código hasta que recibe una solicitud del receptor (en el canal de realimentación) para una retransmisión. En ese punto, el transmisor se detiene, retira la palabra de código particular que el receptor no decodificó correctamente y retransmite la secuencia completa de palabras de código empezando con la corrompida.

En una versión refinada de la ARQ continua conocida como la *ARQ continua con repetición selectiva*, el rendimiento de datos se mejora en forma adicional al retransmitir únicamente la palabra de código que se recibió con errores detectados. En otras palabras, se elimina la necesidad de retransmitir las palabras de código recibidas con éxito que siguen a la palabra de código corrompida.

Los tres tipos de ARQ descritos aquí ofrecen compromisos propios entre la necesidad de un enlace semidúplex o dúplex completo y el requerimiento relativo al uso eficiente de los recursos de comunicación. En cualquier caso, todas ellas se basan en dos premisas:

- La detección de errores, la cual hace que el diseño del decodificador sea relativamente simple.
- El canal de realimentación sin ruido, que no es una restricción severa debido a que la tasa de flujo de información por el canal de realimentación por lo común es bastante baja.

Por estas razones, la ARQ se utiliza ampliamente en los sistemas de comunicación por computadora.

En consecuencia, el hecho de que la FEC requiera únicamente enlaces de una vía para su operación hace que ésta sea una aplicación mucho más extendida que la ARQ. Además, la creciente complejidad de la decodificación de la FEC debida a la necesidad combinada para detección y corrección de errores, ya no es un aspecto de presión práctico, pues el decodificador suele incorporar dentro del mismo un microprocesador o la implementación de VLSI de costo adecuado.

10.2 *Canales sin memoria discretos*

Volviendo al modelo de la figura 10.1*a*, se afirma que el canal de la forma de onda es sin memoria si la salida del detector en un intervalo dado depende únicamente de la señal transmitida en ese intervalo y no de la transmisión previa. Bajo esta condición, es posible modelar la combinación del modulador, el canal de la forma de onda y el detector como un *canal sin memoria discreto*. Un canal de estas características está descrito completamente mediante un conjunto de probabilidades de transición $p(j|i)$, donde i denota un símbolo de entrada del modulador, j un símbolo de salida del demodulador, y $p(j|i)$ la probabilidad de recibir el símbolo j, dado que se envió el símbolo i. (Los canales sin memoria discretos se describieron hasta cierto punto en la sección 9.5.)

El canal sin memoria discreto más simple resulta del uso de símbolos de entrada y salida binarios. Cuando se recurre a la codificación binaria, el modulador sólo tiene los símbolos binarios 0 y 1 como entradas. De igual modo, el decodificador cuenta sólo con entradas binarias si se emplea cuantización binaria de la salida del demodulador; es decir, una *decisión rígida* se lleva a cabo sobre la salida del demodulador, en cuanto a cuál símbolo se transmitió en realidad. En esta situación, tenemos un *canal simétrico binario* (BSC) con *diagrama de probabilidad de transición*, como se muestra en la figura 10.2. El canal simétrico binario, suponiendo un ruido de canal modelado como ruido blanco gaussiano aditivo (AWGN), está descrito por completo por la *probabilidad de transición* p. La mayoría de los sistemas digitales de comunicación codificados emplean codificación binaria con decodificación de decisiones rígida, debido a la simplicidad de la puesta en práctica que ofrece este tipo de métodos. Los *decodificadores de decisión rígida* o *decodificadores*

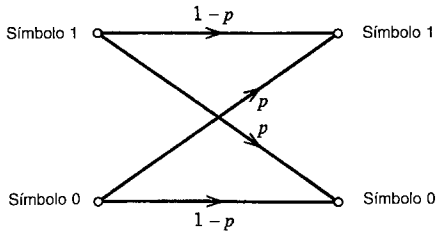


FIGURA 10.2 Diagrama de probabilidad de transición del canal binario simétrico.

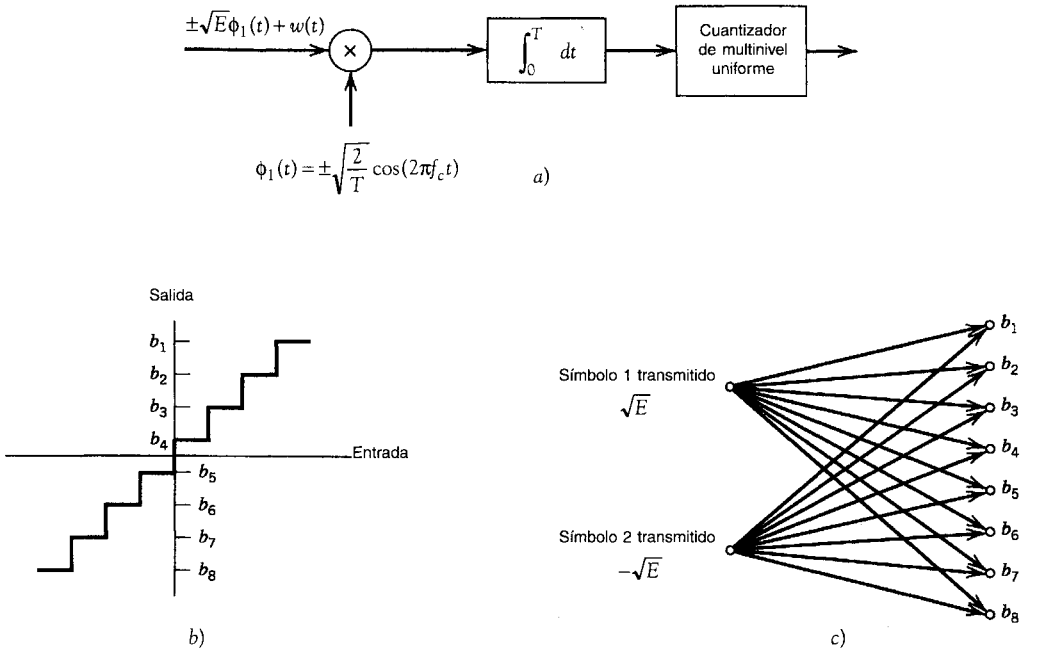


FIGURA 10.3 Canal sin memoria discreto de salida Q-aria y entrada binaria. *a*) Receptor para el corrimiento de fase por llaveo binaria. *b*) Característica de transferencia del cuantizador de multinivel. *c*) Diagrama de la probabilidad de transición del canal. Las partes *b* y *c* se ilustran para ocho niveles de cuantización.

algebraicos, aprovechan la ventaja de la estructura algebraica especial que se incorpora en el diseño de los códigos de canal para conseguir que la decodificación sea relativamente fácil de realizar.

El empleo de decisiones rígidas antes de la decodificación provoca una pérdida irreversible de información en el receptor. Para reducirla se recurre a la codificación de *decisión flexible*, misma que se consigue incluyendo un cuantizador de multinivel a la salida del demodulador, como se ilustra en la figura 10.3*a* para el caso de señales PSK binarias. La característica de entrada-salida del cuantizador se presenta en la figura 10.3*b*. El modulador tiene únicamente los símbolos binarios 0 y 1 como entradas, pero la salida del demodulador tiene ahora un alfabeto con Q símbolos. Suponiendo el uso de un cuantizador, como se describe en la figura 10.3*b*, tenemos $Q = 8$. Un canal de estas características recibe el nombre de *canal sin memoria discreto de salida Q-aria y entrada binaria*. El correspondiente diagrama de probabilidad de transición del canal se muestra en la figura 10.3c. La forma de esta distribución y consecuentemente el desempeño del decodificador, dependen de la ubicación de los niveles de representación del cuantizador, lo cual, a su vez, depende del nivel de la señal y de la varianza del ruido. En consecuencia, el demodulador debe incorporar control de ganancia automático si se va a producir un cuantizador de multinivel efectivo. Además, el uso de decisiones flexibles complica la puesta en práctica del decodificador. A pesar de eso, la decodificación de decisiones flexibles ofrece una mejora importante en el desempeño con respecto a la decodificación de decisiones rígidas al optar por un enfoque probabilístico en vez de algebraico. Por esta razón los decodificadores de decisión flexibles se conocen también como *decodificadores probabilísticos*.

■ REPOSICIÓN DEL TEOREMA DE CODIFICACIÓN DEL CANAL

En el capítulo 9 establecimos el concepto de *capacidad del canal*, el cual, en el caso de un canal sin memoria discreto, representa la cantidad máxima de información transmitida por uso del canal. El

teorema de codificación del canal establece que si un canal sin memoria discreto tiene capacidad C y una fuente genera información a una tasa menor que C, existe entonces una técnica de codificación tal que la salida de la fuente puede transmitirse por el canal con una probabilidad del error de símbolo arbitrariamente baja. Para el caso especial de un canal simétrico binario, el teorema nos indica que si la tasa de código r es menor que la capacidad C del canal, entonces es posible encontrar un código que logre una transmisión sin errores por el canal. Inversamente, no es posible encontrar un código de este tipo si la tasa de código r es mayor que la capacidad C del canal.

El teorema de codificación del canal especifica de ese modo la capacidad C del canal como un *límite fundamental* sobre la tasa a la cual puede llevarse a cabo la transmisión confiable (sin errores) de mensajes por un canal sin memoria discreto. El aspecto que importa no es la relación señal a ruido, mientras que ésta sea suficientemente grande, sino cómo se codifica la entrada del canal.

La característica menos satisfactoria del teorema de codificación del canal es su naturaleza no constructiva. El teorema asegura la existencia de buenos códigos pero no nos dice cómo determinarlos. Por *buenos códigos* queremos decir familias de códigos de canal que son capaces de proporcionar transmisión confiable de información (es decir, a una probabilidad arbitrariamente pequeña del error de símbolo) por un canal ruidoso de interés a tasas de bit hasta un valor máximo menor que la capacidad de ese canal. Las técnicas de codificación de control de errores que se describen en este capítulo ofrecen diferentes métodos de diseño de buenos códigos.

■ Notación

Los códigos que se describen en este capítulo son *binarios*, para los cuales el alfabeto consiste únicamente en los símbolos 0 y 1. En un código de este tipo, las funciones de codificación y decodificación implican operaciones de aritmética binaria de *suma y multiplicación módulo 2* efectuadas sobre palabras de código en el código.

A lo largo de este capítulo utilizamos un signo más ordinario (+) para denotar suma módulo 2. El uso de esta terminología no dará lugar a confusión debido a que todo el capítulo se fundamenta en la aritmética binaria. Al hacerlo de ese modo, evitamos el uso de un símbolo especial $\oplus$, como se hizo en los capítulos anteriores. Por tanto, de acuerdo con la notación que se emplea en este capítulo, las reglas para la suma módulo 2 son como sigue:

$$0 + 0 = 0$$
$$1 + 0 = 0$$
$$0 + 1 = 1$$
$$1 + 1 = 0$$

Puesto que $1 + 1 = 0$, se concluye que $1 = -1$. De ese modo, en la aritmética binaria, la sustracción es la misma que la adición. Las reglas para la multiplicación módulo 2 son las siguientes:

$$0 \times 0 = 0$$
$$1 \times 0 = 0$$
$$0 \times 1 = 0$$
$$1 \times 1 = 1$$

La división es trivial en cuanto a que tenemos

$$1 \div 1 = 1$$
$$0 \div 1 = 0$$

y la división entre 0 no está permitida. La suma módulo 2 es la operación OR EXCLUSIVA en la lógica, y la multiplicación módulo 2 es la operación AND.

10.3 *Códigos de bloques lineales*

Se dice que un código es lineal *si cualesquiera dos palabras de código en el código pueden sumarse en aritmética módulo 2 para producir una tercera palabra de código en el código*. Considere entonces un código de bloque lineal (n, k), en el cual k bits de los n bits de código son siempre idénticos a la secuencia de mensaje que se va a transmitir. Los $n - k$ bits en la porción restante se calculan a partir de los bits de mensaje de acuerdo con la regla de codificación preestablecida que determina la estructura matemática del código. Por tanto, estos $n - k$ bits se conocen como *bits de verificación de paridad generalizados* o simplemente *bits de paridad*. Los códigos de bloque en los cuales se transmiten los bits de mensaje en una forma inalterada se denominan *códigos sistemáticos*. Para las aplicaciones que requieren *tanto* detección *como* corrección de errores, el uso de códigos de bloque sistemáticos simplifica la puesta en práctica del decodificador.

Considere que $m_0, m_1, \ldots, m_{k-1}$ constituyen un bloque de k bits de mensaje arbitrarios. De ese modo tenemos 2^k bloques de mensaje distintos. Suponga que esta secuencia de bits de mensaje se aplica a un codificador de bloques lineal, produciéndose una palabra de código de n bits cuyos elementos se denotan por medio de $c_0, c_1, \ldots, c_{n-1}$. Sean $b_0, b_1, \ldots, b_{n-k-1}$ los $(n - k)$ bits de paridad en la palabra de código. Para que el código posea una estructura sistemática, una palabra de código se divide en dos partes, una de las cuales la ocupan los bits de mensaje y la otra los bits de paridad. Claramente, tenemos la opción de enviar los bits de mensaje de una palabra de código antes de los bits de paridad o viceversa. La primera opción se ilustra en la figura 10.4 y su uso se describe en lo que sigue.

De acuerdo con la representación de la figura 10.4, los $(n - k)$ bits más a la izquierda de una palabra de código son idénticos a los bits de paridad correspondientes, y los k bits más a la derecha de la palabra de código son idénticos a los bits de mensaje correspondientes. Por tanto, es posible escribir

$$c_i = \begin{cases} b_i, & i = 0, 1, \ldots, n-k-1 \\ m_{i+k-n}, & i = n-k, n-k+1, \ldots, n-1 \end{cases} \tag{10.1}$$

Los $(n - k)$ bits de paridad son *sumas lineales* de los k bits de mensaje, como se indica mediante la relación generalizada

$$b_i = p_{0i}m_0 + p_{1i}m_1 + \ldots + p_{k-1,i}m_{k-1} \tag{10.2}$$

donde los coeficientes se definen del modo que sigue:

$$p_{ij} = \begin{cases} 1 & \text{si } b_i \text{ depende de } m_j \\ 0 & \text{en otro caso} \end{cases} \tag{10.3}$$

Los coeficientes p_{ij} se eligen de tal manera que los renglones de la matriz generadora sean linealmente independientes y las ecuaciones de paridad sean *únicas*.

El sistema de las ecuaciones (10.1) y (10.2) define la estructura matemática del código de bloques lineal (n, k). Este sistema de ecuaciones puede escribirse en una forma compacta utili-

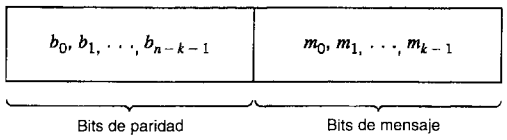


FIGURA 10.4 Estructura de la palabra de código sistemática.

zando la notación de matrices. Para proseguir con esta reformulación, definimos el *vector de mensaje* de 1 por k, o *vector de información* $\mathbf{m}$, el vector de paridad $\mathbf{b}$ de 1 por $(n-k)$ y el vector $\mathbf{c}$ de 1 por n de la manera siguiente:

$$\mathbf{m} = [m_0, m_1, \ldots, m_{k-1}] \tag{10.4}$$

$$\mathbf{b} = [b_0, b_1, \ldots, b_{n-k-1}] \tag{10.5}$$

$$\mathbf{c} = [c_0, c_1, \ldots, c_{n-1}] \tag{10.6}$$

Advierta que los tres son *vectores renglón*. Se adopta el empleo de vectores renglón en este capítulo en aras de ser consistentes con la notación que se utiliza comúnmente en la literatura de la codificación. De ese modo podemos reescribir el conjunto de ecuaciones simultáneas definiendo los bits de paridad en la forma de matriz compacta:

$$\mathbf{b} = \mathbf{mP} \tag{10.7}$$

donde $\mathbf{P}$ es la *matriz de coeficientes* de k por $(n-k)$ definida por

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0,n-k-1} \\ p_{10} & p_{11} & \cdots & p_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ p_{k-1,0} & p_{k-1,1} & \cdots & p_{k-1,n-k-1} \end{bmatrix} \tag{10.8}$$

donde p_{ij} es 0 o 1.

A partir de las definiciones dadas en las ecuaciones (10.4)-(10.6), vemos que es posible expresar $\mathbf{c}$ como un vector renglón dividido en términos de los vectores $\mathbf{m}$ y $\mathbf{b}$ en la forma siguiente:

$$\mathbf{c} = [\mathbf{b} \vdots \mathbf{m}] \tag{10.9}$$

Por tanto, al sustituir la ecuación (10.7) en la (10.9) y factorizar el vector de mensaje común $\mathbf{m}$, obtenemos

$$\mathbf{c} = \mathbf{m}[\mathbf{P} \vdots \mathbf{I}_k] \tag{10.10}$$

donde $\mathbf{I}_k$ es la *matriz identidad* de k por k:

$$\mathbf{I}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{10.11}$$

Defina la *matriz generadora* de k por n

$$\mathbf{G} = [\mathbf{P} \vdots \mathbf{I}_k] \tag{10.12}$$

La matriz $\mathbf{G}$ generadora de la ecuación (10.12) se dice que está en *forma canónica* porque sus k renglones son linealmente independientes; es decir, no es posible expresar ningún renglón de la matriz $\mathbf{G}$ como una combinación lineal de los renglones restantes. Utilizando la definición de la matriz $\mathbf{G}$ generadora, podemos simplificar la ecuación (10.10) como

$$\mathbf{c} = \mathbf{mG} \tag{10.13}$$

El conjunto completo de palabras de código, referido simplemente como *el código*, se genera de acuerdo con la ecuación (10.13) dejando que el vector de mensaje $\mathbf{m}$ varíe a través del conjunto de todas las k–tupletas (1 por k vectores) binarias. Además la suma de cualesquiera dos palabras de código es otra palabra de código. Esta propiedad básica de los códigos de bloque

lineales se llama *cerradura*. Para probar su validez, considere un par de vectores de código $\mathbf{c}_i$ y $\mathbf{c}_j$ correspondientes a un par de vectores de mensaje $\mathbf{m}_i$ y $\mathbf{m}_j$, respectivamente. Al utilizar la ecuación (10.13), es posible expresar la suma de $\mathbf{c}_i$ y $\mathbf{c}_j$ como

$$\begin{aligned}\mathbf{c}_i + \mathbf{c}_j &= \mathbf{m}_i\mathbf{G} + \mathbf{m}_j\mathbf{G} \\ &= (\mathbf{m}_i + \mathbf{m}_j)\mathbf{G}\end{aligned}$$

La suma módulo 2 de $\mathbf{m}_i$ y $\mathbf{m}_j$ representa un nuevo vector de mensaje. De modo correspondiente, la suma módulo 2 de $\mathbf{c}_i$ y $\mathbf{c}_j$ representa un nuevo vector de código.

Hay otra manera de expresar la relación entre los bits de mensaje y los bits de verificación de paridad de un código de bloques lineal. Considere que $\mathbf{H}$ denota una matriz de $(n - k)$ por n, definida como

$$\mathbf{H} = [\mathbf{I}_{n-k} \vdots \mathbf{P}^T] \tag{10.14}$$

donde $\mathbf{P}^T$ es una matriz de $(n - k)$ por k, que representa a la transpuesta de la matriz de coeficientes $\mathbf{P}$, e $\mathbf{I}_{n-k}$ es la matriz identidad de $(n - k)$ por $(n - k)$. En consecuencia, es viable realizar la siguiente multiplicación de las matrices particionadas:

$$\begin{aligned}\mathbf{H}\mathbf{G}^T &= [\mathbf{I}_{n-k} \vdots \mathbf{P}^T] \begin{bmatrix} \mathbf{P}^T \\ \cdots \\ \mathbf{I}_k \end{bmatrix} \\ &= \mathbf{P}^T + \mathbf{P}^T\end{aligned}$$

donde hemos recurrido al hecho de que la multiplicación de una matriz rectangular por una matriz identidad de dimensiones compatibles deja a la matriz sin cambio. En aritmética módulo 2 tenemos que $\mathbf{P}^T + \mathbf{P}^T = \mathbf{0}$, donde $\mathbf{0}$ denota una matriz nula de $(n - k)$ por k (es decir, una matriz que tiene ceros en todos sus elementos). Por tanto,

$$\mathbf{H}\mathbf{G}^T = \mathbf{0} \tag{10.15}$$

De manera equivalente, tenemos que $\mathbf{G}\mathbf{H}^T = \mathbf{0}$, donde cero es una nueva matriz nula. Al multiplicar después ambos miembros de la ecuación (10.13) por $\mathbf{H}^T$, la transpuesta de $\mathbf{H}$, y al utilizar la ecuación (10.15), se obtiene

$$\begin{aligned}\mathbf{c}\mathbf{H}^T &= \mathbf{m}\mathbf{G}\mathbf{H}^T \\ &= \mathbf{0}\end{aligned} \tag{10.16}$$

$\mathbf{H}$ se conoce como la *matriz de verificación de paridad* del código, y al conjunto especificado por la ecuación (10.16) se asigna el nombre de *ecuaciones de verificación de paridad*.

La ecuación (10.13) generadora y la ecuación (10.16) del detector de verificación de paridad resultan básicas para la descripción y operación de un código de bloques lineal. Estas dos ecuaciones se describen en la forma de diagramas de bloque en las figuras 10.5*a* y 10.5*b*, respectivamente.

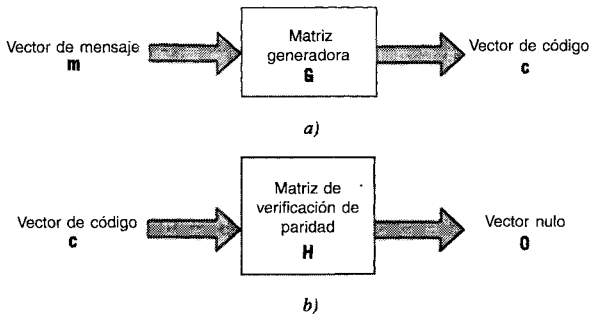


FIGURA 10.5 Representaciones de diagrama de bloques de la ecuación genérica (10.13) y la ecuación de verificación de paridad.

▶ EJEMPLO 10.1 Códigos de repetición

Los *códigos de repetición* representan el tipo de códigos de bloques lineales más simple. En particular, un solo bit de mensaje se codifica en un bloque de n bits idénticos, produciendo un código de bloque $(n, 1)$. Un código de tales características permite la provisión de una cantidad variable de redundancia. Sólo existen dos palabras de código en el código: una palabra de código de puros ceros y una palabra de código de puros unos.

Considere, por ejemplo, el caso de un código de repetición con $k = 1$ y $n = 5$. En este caso, tenemos cuatro bits de paridad que son iguales al bit de mensaje. Por tanto, la matriz identidad $\mathbf{I}_k = 1$ y la matriz de coeficientes $\mathbf{P}$ están compuestas por un vector de 1 por 4 cuya totalidad de elementos son 1. De modo correspondiente, la matriz generadora es igual a un vector renglón de puros unos, como se indica mediante

$$\mathbf{G} = [1\ 1\ 1\ 1 \vdots 1]$$

La transpuesta de la matriz de coeficientes $\mathbf{P}$, a saber, la matriz $\mathbf{P}^T$, consiste en un vector de 4 por 1 que tiene al 1 como el valor de todos sus elementos. La matriz identidad $\mathbf{I}_{n-k}$ está compuesta por una matriz de 4 por 4. Por tanto, la matriz de verificación de paridad es igual a

$$\mathbf{H} = \begin{bmatrix} 1\ 0\ 0\ 0 \vdots 1 \\ 0\ 1\ 0\ 0 \vdots 1 \\ 0\ 0\ 1\ 0 \vdots 1 \\ 0\ 0\ 0\ 1 \vdots 1 \end{bmatrix}$$

Puesto que el vector de mensaje consiste en un símbolo binario simple, 0 o 1, se concluye a partir de la ecuación (10.13) que sólo hay dos palabras de código: 00000 y 11111 en el código de repetición (5, 1), como se esperaba. Advierta también que $\mathbf{HG}^T = \mathbf{0}$, módulo 2, en concordancia con la ecuación (10.15). ◀

■ SÍNDROME: DEFINICIÓN Y PROPIEDADES

La matriz $\mathbf{G}$ generadora se usa en la operación de codificación en el transmisor. Por otra parte, la matriz $\mathbf{H}$ de verificación de paridad se usa en la operación de decodificación en el receptor. En el contexto de la última operación, considere que $\mathbf{r}$ denota el *vector recibido* de 1 por n que resulta del envío del vector de código $\mathbf{c}$ por un canal ruidoso. Expresamos el vector $\mathbf{r}$ como la suma del vector de código original $\mathbf{c}$ y de un vector $\mathbf{e}$:

$$\mathbf{r} = \mathbf{c} + \mathbf{e} \tag{10.17}$$

El vector $\mathbf{e}$ se denomina *vector de error* o *patrón de error*. El i-ésimo elemento de $\mathbf{e}$ es igual a 0 si el elemento correspondiente de $\mathbf{r}$ es el mismo que el de $\mathbf{c}$. Por otra parte, el i-ésimo elemento de $\mathbf{e}$ es igual a 1 si el elemento correspondiente de $\mathbf{r}$ es diferente del de $\mathbf{c}$, en cuyo caso se dice que ha ocurrido un error en la i-ésima posición. Es decir, para $i = 1, 2, \ldots, n$, tenemos

$$e_i = \begin{cases} 1 & \text{si un error ha ocurrido en la } i\text{-ésima posición} \\ 0 & \text{en otro caso} \end{cases} \tag{10.18}$$

El receptor tiene la tarea de decodificar el vector de código $\mathbf{c}$ a partir del vector recibido $\mathbf{r}$. El algoritmo que se usa comúnmente para efectuar esta operación de decodificado empieza con el cálculo de un vector de 1 por $(n - k)$ llamado *vector del síndrome de error* o simplemente *síndrome*.[3] La importancia de este síndrome radica en que depende exclusivamente del patrón de error.

Dado un vector recibido $\mathbf{r}$ de 1 por n, el síndrome correspondiente se define de manera formal como

$$\mathbf{s} = \mathbf{r}\mathbf{H}^T \tag{10.19}$$

En consecuencia, el síndrome tiene las siguientes propiedades importantes.

Propiedad 1

El síndrome depende únicamente del patrón de error y no de la palabra de código transmitida.

Para demostrar esta propiedad, recurrimos primero a las ecuaciones (10.17), (10.19) y después a la (10.16) para obtener

$$\begin{aligned}\mathbf{s} &= (\mathbf{c}+\mathbf{e})\mathbf{H}^T \\ &= \mathbf{c}\mathbf{H}^T + \mathbf{e}\mathbf{H}^T \\ &= \mathbf{e}\mathbf{H}^T\end{aligned} \tag{10.20}$$

Por tanto, la matriz **H** de verificación de paridad de un código nos permite calcular el síndrome **s**, mismo que depende solamente del patrón de errores **e**.

Propiedad 2

Todos los patrones de errores que difieren en una palabra de código tienen el mismo síndrome.

Para k bits de mensaje, hay 2^k vectores de código distintos denotados como $\mathbf{c}_i$, $i = 0, 1, ..., 2^k - 1$. De manera correspondiente, para cualquier patrón de errores **e**, definimos los 2^k vectores distintos $\mathbf{e}_i$ como

$$\mathbf{e}_i = \mathbf{e} + \mathbf{c}_i, \quad i = 0, 1, \ldots, 2^k - 1 \tag{10.21}$$

El conjunto de vectores $\{\mathbf{e}_i, i = 0, 1, ..., 2^k - 1\}$ así definido se denomina un *coconjunto del código*. En otras palabras, un coconjunto tiene exactamente 2^k elementos que difieren a lo más por un vector de código. De esa manera, un código de bloques lineal (n, k) tiene 2^{n-k} posibles coconjuntos. En cualquier caso, al multiplicar ambos miembros de la ecuación (10.21) por la matriz $\mathbf{H}^T$, obtenemos

$$\begin{aligned}\mathbf{e}_i\mathbf{H}^T &= \mathbf{e}\mathbf{H}^T + \mathbf{c}_i\mathbf{H}^T \\ &= \mathbf{e}\mathbf{H}^T\end{aligned} \tag{10.22}$$

que es independiente del índice i. Por tanto, es posible enunciar que cada coconjunto del código está caracterizado por un síndrome único.

Podemos poner en perspectiva las propiedades 1 y 2 al expandir la ecuación (10.20). Específicamente, con la matriz **H** teniendo la forma sistemática dada en la ecuación (10.14), donde la propia matriz **P** se define a partir de la ecuación (10.8), encontramos con base en la ecuación (10.20) que los $(n - k)$ elementos del símbolo **s** son combinaciones lineales de los n elementos del patrón de errores **e**, en la forma que indica

$$\begin{aligned}s_0 &= e_0 + e_{n-k}p_{00} + e_{n-k+1}p_{10} + \cdots + e_{n-1}p_{k-1,0} \\ s_1 &= e_1 + e_{n-k}p_{01} + e_{n-k+1}p_{11} + \cdots + e_{n-1}p_{k-1,1} \\ &\vdots \\ s_{n-k-1} &= e_{n-k-1} + e_{n-k}p_{0,n-k-1} + \cdots + e_{n-1}p_{k-1,n-k-1}\end{aligned} \tag{10.23}$$

Este conjunto de $(n - k)$ ecuaciones lineales muestra claramente que el síndrome contiene información acerca de los patrones de errores y que, por tanto, es posible utilizarla para la detección de los mismos. Sin embargo, debe advertirse que el conjunto de ecuaciones está *indeterminado* en el sentido de que tenemos más incógnitas que ecuaciones. Por tanto, *no* hay una solución única para el patrón de errores; más bien, existen 2^n patrones de errores que satisfacen la ecuación (10.23) y, en consecuencia, producen el mismo síndrome, de acuerdo con la propiedad 2 y la ecuación (10.22). En particular, con 2^{n-k} posibles vectores de síndrome, la información contenida en el síndrome **s** acerca del patrón de errores **e** *no* es suficiente para que el decodificador calcule el valor exacto del vector de código transmitido. A pesar de eso, el conocimiento del síndrome **s** reduce la búsqueda del verdadero patrón de errores **e** de 2^n a 2^{n-k}

posibilidades. Dadas estas posibilidades, el decodificador tiene la tarea de realizar la mejor selección a partir de los coconjuntos correspondientes a **s**.

Consideraciones de la distancia mínima

Considere un par de vectores de código $\mathbf{c}_1$ y $\mathbf{c}_2$ que tienen el mismo número de elementos. La *distancia de Hamming* $d(\mathbf{c}_1, \mathbf{c}_2)$ entre tal par de vectores de código se define como el número de localidades en las cuales difieren sus elementos respectivos.

El *peso de Hamming* $w(\mathbf{c})$ de un vector de código $\mathbf{c}$ se define como el número de elementos distintos de cero en el vector de código. De manera equivalente, es posible establecer que el peso de Hamming de un vector de código corresponde a la distancia entre el vector de código y el vector de código de puros ceros.

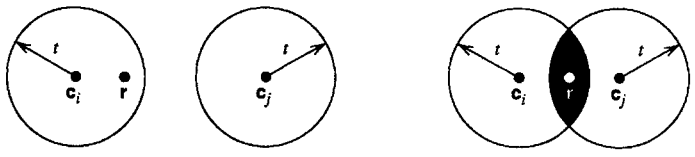
La *distancia mínima* $d_{\text{mín}}$ de un código de bloques lineal se define como la distancia de Hamming más pequeña entre cualquier par de vectores de código en el código. Es decir, la distancia mínima es la misma que el peso de Hamming más pequeño de la diferencia entre cualquier par de vectores de código. A partir de la propiedad de cerradura de los códigos de bloques lineales, la suma (o diferencia) de dos vectores de código es otro vector de código. Por tanto, es posible establecer *que la distancia mínima de un código de bloques lineal es el peso de Hamming más pequeño de los vectores de código distintos de cero en el código.*

La distancia mínima $d_{\text{mín}}$ se relaciona con la estructura de la matriz $\mathbf{H}$ de verificación de paridad del código de una manera fundamental. Con base en la ecuación (10.16) sabemos que un código de bloques lineal se define por medio del conjunto de todos los vectores de código para los cuales $\mathbf{cH}^T = \mathbf{0}$, donde $\mathbf{H}^T$ es la transpuesta de la matriz $\mathbf{H}$ de verificación de paridad. Dejemos que la matriz $\mathbf{H}$ se exprese en términos de sus columnas de la manera siguiente:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n] \tag{10.24}$$

En ese caso, para que un vector de código $\mathbf{c}$ satisfaga la condición $\mathbf{cH}^T = \mathbf{0}$, el vector $\mathbf{c}$ debe tener unos en posiciones tales que los renglones correspondientes de $\mathbf{H}^T$ sumen el vector cero. Sin embargo, por definición, el número de unos en un vector de código es el peso de Hamming de dicho vector. Asimismo, el peso de Hamming más pequeño de los vectores de código distintos de cero en un código de bloques lineal es igual a la distancia mínima del código. Por tanto, *la distancia mínima de un código de bloques lineal se define mediante el número mínimo de renglones de la matriz* $\mathbf{H}^T$ cuya suma es igual al vector cero.

La distancia mínima de un código de bloques lineal, $d_{\text{mín}}$, es un parámetro importante del código. Específicamente, determina la capacidad de corrección de errores de este último. Suponga que se requiere un código de bloque lineal (n, k) para detectar y corregir todos los patrones de errores (sobre un canal simétrico binario) y cuyo peso de Hamming es menor o igual que t. Es decir, si se transmite un vector de código $\mathbf{c}_i$ en el código y el vector recibido es $\mathbf{r} = \mathbf{c}_i + \mathbf{e}$, requerimos que la salida del decodificador $\hat{\mathbf{c}} = \mathbf{c}_i$, siempre que el patrón de errores $\mathbf{e}$ tenga un peso de Hamming $w(\mathbf{e}) \leq t$. Suponemos que los 2^k vectores de código en el código se transmiten con igual probabilidad. La mejor estrategia para el decodificador es entonces recoger el vector de código más cercano al vector recibido $\mathbf{r}$, es decir, aquel para el cual la distancia de Hamming $d(\mathbf{c}_i, \mathbf{r})$ es la más pequeña. Con este tipo de estrategia, el decodificador será capaz de detectar y corregir todos los patrones de errores del peso de Hamming $w(\mathbf{e}) \leq t$, siempre que la distancia mínima del código sea igual o mayor que $2t + 1$. Es posible demostrar la validez de este requerimiento adoptando una interpretación geométrica del problema. En particular, los vectores de código de 1 por n y los vectores recibidos de 1 por n se representan como puntos en un espacio n–dimensional. Suponga que construimos dos esferas, cada una de radio t, alrededor de los puntos que representan a los vectores de código $\mathbf{c}_i$ y $\mathbf{c}_j$. Considere que estas dos esferas son disjuntas, como se indica en la figura 10.6*a*. Para que esta condición se cumpla, requerimos que $d(\mathbf{c}_i, \mathbf{c}_j) \geq 2t + 1$. Si luego el vector de código $\mathbf{c}_i$ se transmite y la distancia de Hamming



a) *b)*

FIGURA 10.6 *a)* Distancia de Hamming $d(\mathbf{c}_i, \mathbf{c}_j) \geq 2t + 1$. *b)* Distancia de Hamming $d(\mathbf{c}_i, \mathbf{c}_j) < 2t$. El vector recibido se denota mediante **r**.

$d(\mathbf{c}_i, \mathbf{r}) \leq t$, es claro que el decodificador escogerá $\mathbf{c}_i$ cuando éste sea el vector de código más cercano al vector recibido **r**. Si, por otro lado, la distancia de Hamming $d(\mathbf{c}_i, \mathbf{c}_j) \leq 2t$, se interceptan las dos esferas en torno a $\mathbf{c}_i$ y $\mathbf{c}_j$, como se describe en la figura 10.6*b*. Aquí vemos que si se transmite $\mathbf{c}_i$, ahí existe un vector recibido **r** tal que la distancia de Hamming $d(\mathbf{c}_i, \mathbf{r}) \leq t$ y aun **r** es tan cercano a $\mathbf{c}_j$ como lo es a $\mathbf{c}_i$. Claramente, existe ahora la posibilidad de que el decodificador escoja el vector $\mathbf{c}_j$, lo cual es incorrecto. En consecuencia, concluimos que *un código de bloques lineal (n, k) tiene la capacidad de corregir todos los patrones de errores de peso t o menos si, y sólo si,*

$$d(\mathbf{c}_i, \mathbf{c}_j) \geq 2t+1 \quad \text{para toda } \mathbf{c}_i \text{ y } \mathbf{c}_j$$

Sin embargo, por definición, la distancia más pequeña entre cualquier par de vectores de código en un código corresponde a la distancia mínima del código $d_{\text{mín}}$. Por tanto, es posible establecer que *un código de bloques lineal (n, k) de distancia mínima* $d_{\text{mín}}$ *puede corregir hasta t errores si, y sólo si,*

$$t \leq \left\lfloor \tfrac{1}{2}(d_{\text{mín}} - 1) \right\rfloor \tag{10.25}$$

donde $\lfloor\ \rfloor$ *denota el entero mayor*, menor o igual que la cantidad encerrada. La ecuación (10.25) da un sentido cuantitativo a la capacidad de corrección de errores de un código de bloques lineal.

■ DECODIFICACIÓN DEL SÍNDROME

Estamos listos ahora para describir un esquema de decodificación basado en el síndrome para códigos de bloque lineales. Sean $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{2^k}$ los 2^k vectores de código de un código de bloques lineal (n, k). Sea **r** el vector recibido, el cual puede tener uno de 2^n valores posibles. El receptor tiene la tarea de dividir los 2^n vectores recibidos posibles en 2^k subconjuntos disjuntos $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_{2^k}$, de tal forma que el i-ésimo subconjunto $\mathscr{D}_i$ corresponda al vector de código $\mathbf{c}_i$ para $1 \leq i \leq 2^k$. El vector recibido **r** se decodifica en $\mathbf{c}_i$ si éste se encuentra en el i-ésimo subconjunto. Para que la decodificación sea correcta, **r** debe estar en el subconjunto que pertenece al vector de código $\mathbf{c}_i$ que realmente fue enviado.

Los 2^k subconjuntos descritos aquí constituyen un *arreglo estándar* del código de bloques lineal. Para construirlo, aprovechamos la estructura lineal del código procediendo del modo siguiente:

1. Los 2^k vectores de código se colocan en un renglón con el vector de código de puros ceros $\mathbf{c}_1$ como el elemento más a la izquierda.
2. Se escoge un patrón de errores $\mathbf{e}_2$, se coloca bajo $\mathbf{c}_1$ y se forma un segundo renglón al agregar $\mathbf{e}_2$ a cada uno de los vectores de código restantes en el primer renglón; es importante que el patrón de errores elegidos como primer elemento en un renglón no haya aparecido previamente en el arreglo estándar.
3. El paso 2 se repite hasta que todos los patrones de error posibles se hayan tomado en cuenta.

La figura 10.7 ilustra la estructura de un arreglo estándar construido de esa manera. Las 2^k columnas de este arreglo representan a los subconjuntos disjuntos $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_{2^k}$. Los 2^{n-k} renglones del arreglo representan los coconjuntos del código, y sus primeros elementos $\mathbf{e}_2, \ldots, \mathbf{e}_{2^{n-k}}$ reciben el nombre de *guías de coconjunto*.

$$
\begin{array}{ccccccc}
\mathbf{c}_1 = 0 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_i & \cdots & \mathbf{c}_{2^k} \\
\mathbf{e}_2 & \mathbf{c}_2 + \mathbf{e}_2 & \mathbf{c}_3 + \mathbf{e}_2 & \cdots & \mathbf{c}_i + \mathbf{e}_2 & \cdots & \mathbf{c}_{2^k} + \mathbf{e}_2 \\
\mathbf{e}_3 & \mathbf{c}_2 + \mathbf{e}_3 & \mathbf{c}_3 + \mathbf{e}_3 & \cdots & \mathbf{c}_i + \mathbf{e}_3 & \cdots & \mathbf{c}_{2^k} + \mathbf{e}_3 \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\mathbf{e}_j & \mathbf{c}_2 + \mathbf{e}_j & \mathbf{c}_3 + \mathbf{e}_j & \cdots & \mathbf{c}_i + \mathbf{e}_j & \cdots & \mathbf{c}_{2^k} + \mathbf{e}_j \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\mathbf{e}_{2^{n-k}} & \mathbf{c}_2 + \mathbf{e}_{2^{n-k}} & \mathbf{c}_3 + \mathbf{e}_{2^{n-k}} & & \mathbf{c}_i + \mathbf{e}_{2^{n-k}} & & \mathbf{c}_{2^k} + \mathbf{e}_{2^{n-k}}
\end{array}
$$

FIGURA 10.7 Arreglo estándar de un código de bloques (n, k).

Para un canal dado, la probabilidad de error de decodificación se minimiza cuando se eligen los patrones de error más probables (es decir, aquellos con la probabilidad de ocurrencia más grande) como las guías de coconjuntos. En el caso de un canal simétrico binario, cuanto más pequeño es el peso de Hamming de un patrón de errores, tanto más probable es que éste ocurra. Por tanto, el arreglo estándar debe construirse con cada guía de coconjunto teniendo el peso de Hamming mínimo en su coconjunto.

Después de esto es posible describir un procedimiento de decodificación para un código de bloques lineal:

1. Para el vector recibido $\mathbf{r}$, calcule el síndrome $\mathbf{s} = \mathbf{r}\mathbf{H}^T$.
2. Dentro del coconjunto caracterizado por el síndrome $\mathbf{s}$, identifique la guía de coconjunto (es decir, el patrón de errores con la probabilidad de ocurrencia más alta); llámelo $\mathbf{e}_0$.
3. Calcule el vector de código

$$\mathbf{c} = \mathbf{r} + \mathbf{e}_0 \tag{10.26}$$

como la versión decodificada del vector recibido $\mathbf{r}$.

Este procedimiento recibe el nombre de *decodificación del síndrome*.

▶ EJEMPLO 10.2 Códigos de Hamming[4]

Considere una familia de códigos de bloque lineales (n, k) que tienen los siguientes parámetros:

Longitud de bloque:	$n = 2^m - 1$
Número de bits de mensaje:	$k = 2^m - m - 1$
Número de bits de paridad:	$n - k = m$

donde $m \geq 3$. Éstos son los llamados códigos de Hamming.

Considere, por ejemplo, el código de Hamming (7, 4) con $n = 7$ y $k = 4$, correspondiendo a $m = 3$. La matriz generadora del código debe tener una estructura que se apegue a la ecuación (10.12). La siguiente matriz representa una matriz generadora apropiada para el código de Hamming (7, 4):

$$
\mathbf{G} = \left[\begin{array}{ccc:cccc}
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}\right]
= \left[\underbrace{\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{array}}_{\mathbf{P}} \;\; \underbrace{\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}}_{\mathbf{I}_k}\right]
$$

TABLA 10.1 *Palabras de código de un código de Hamming (7, 4)*

Palabra de mensaje	*Palabra de código*	*Peso de una palabra de código*	*Palabra de mensaje*	*Palabra de código*	*Peso de una palabra de código*
0 0 0 0	0 0 0 0 0 0 0	0	1 0 0 0	1 1 0 1 0 0 0	3
0 0 0 1	1 0 1 0 0 0 1	3	1 0 0 1	0 1 1 1 0 0 1	4
0 0 1 0	1 1 1 0 0 1 0	4	1 0 1 0	0 0 1 1 0 1 0	3
0 0 1 1	0 1 0 0 0 1 1	3	1 0 1 1	1 0 0 1 0 1 1	4
0 1 0 0	0 1 1 0 1 0 0	3	1 1 0 0	1 0 1 1 1 0 0	4
0 1 0 1	1 1 0 0 1 0 1	4	1 1 0 1	0 0 0 1 1 0 1	3
0 1 1 0	1 0 0 0 1 1 0	3	1 1 1 0	0 1 0 1 1 1 0	4
0 1 1 1	0 0 1 0 1 1 1	4	1 1 1 1	1 1 1 1 1 1 1	7

La correspondiente matriz de verificación de paridad está dada por

$$\mathbf{H} = \underbrace{\left[\begin{array}{ccc:cccc} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array}\right]}_{\mathbf{I}_{n-k}} \quad \underbrace{\phantom{\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]}}_{\mathbf{P}^T}$$

Con $k = 4$ hay $2^k = 16$ palabras de mensaje distintas, las cuales se listan en la tabla 10.1. Para una palabra de mensaje determinada, la palabra de código correspondiente se obtiene utilizando la ecuación (10.13). Por tanto, la aplicación de dicha ecuación produce las 16 palabras de código que se listan en la tabla 10.1.

En la tabla 10.1 hemos listado también los pesos de Hamming de las palabras de código individuales en el código de Hamming (7, 4). Puesto que el más pequeño de los pesos de Hamming para las palabras de código distintas de cero es 3, se sigue que la distancia mínima del código es 3. En realidad, los códigos de Hamming tienen la propiedad de que la distancia mínima $d_{\text{mín}} = 3$, independientemente del valor asignado al número de bits de paridad m.

Para ilustrar la relación entre la distancia mínima $d_{\text{mín}}$ y la estructura de la matriz **H** de verificación de paridad, considere la palabra de código 0110100. En la multiplicación de matrices definida por la ecuación (10.16), los elementos distintos de cero de esta palabra de código se "extraen" fuera de la segunda, tercera y quinta columnas de la matriz **H**, produciéndose

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Podemos efectuar cálculos similares para las 14 palabras de código distintas de cero restantes. De ese modo encontramos que el número más pequeño de columnas en **H** que suma cero es 3, confirmando el enunciado anterior de que $d_{\text{mín}} = 3$.

Una propiedad importante de los códigos de Hamming es que satisfacen la condición de la ecuación (10.25) con el signo de igualdad, suponiendo que $t = 1$. Esto significa que los códigos de Hamming son *códigos perfectos binarios de corrección de errores simples*.

Suponiendo patrones de errores simples, es posible formular las siete guías de coconjunto que se listan en la columna del lado derecho de la tabla 10.2. Los correspondientes 2^3 síndromes, listados en la columna de la izquierda, se calculan de acuerdo con la ecuación (10.20). El síndrome cero significa que no hay errores de transmisión.

Suponga, por ejemplo, que se envía el vector de código [1110010] y que el vector recibido

TABLA 10.2 *Tabla de decodificación para el código de Hamming (7, 4) definido en la tabla 10.1*

Síndrome	*Patrón de errores*
000	0000000
100	1000000
010	0100000
001	0010000
110	0001000
011	0000100
111	0000010
101	0000001

corresponde a [1100010] con un error en el tercer bit. Utilizando la ecuación (10.19), se calcula que el síndrome es

$$\mathbf{s} = [1100010]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad 1]$$

De acuerdo con la tabla 10.2, la guía de coconjunto correspondiente (es decir, el patrón de errores con la probabilidad de ocurrencia más alta) se encuentra que es [0010000], lo que indica correctamente que el tercer bit del vector recibido es erróneo. De esa manera, agregando este patrón de errores al vector recibido, de acuerdo con la ecuación (10.26), se produce el vector de código correcto realmente enviado. ◀

CÓDIGO DUAL

Dado un código de bloques lineal, es posible definir su *dual* de la manera siguiente. Al tomar la transpuesta en ambos miembros de la ecuación (10.15), tenemos

$$\mathbf{G}\mathbf{H}^T = \mathbf{0}$$

donde $\mathbf{H}^T$ es la transpuesta de la matriz de verificación de paridad del código y $\mathbf{0}$ es una nueva matriz cero. Esta ecuación sugiere que todo código de bloques lineal (n, k) con matriz $\mathbf{G}$ generadora y matriz $\mathbf{H}$ de verificación de paridad tiene un *código dual* con parámetros $(n, n - k)$, matriz generadora $\mathbf{H}$ y matriz de verificación de paridad $\mathbf{G}$.

10.4 *Códigos cíclicos*

Los códigos cíclicos forman una subclase de códigos de bloques lineales. En realidad, muchos de los códigos de bloques lineales importantes descubiertos hasta la fecha son cíclicos o bastante relacionados con este tipo de códigos. Una ventaja de los códigos cíclicos sobre la mayor

parte de los otros tipos de códigos es que son más fáciles de codificar. Además, los códigos cíclicos poseen una estructura matemática perfectamente definida, la cual ha conducido al desarrollo de esquemas de decodificación muy eficientes para ellos.

Se afirma que un código binario será *cíclico* si exhibe dos propiedades fundamentales:

1. *Propiedad de linealidad*: la suma de cualesquiera dos palabras de código en el código es también una palabra de código.
2. *Propiedad cíclica*: cualquier corrimiento cíclico de una palabra de código en el código también es una palabra de código.

La propiedad 1 reenuncia el hecho de que el código cíclico es un código de bloques lineal (es decir, puede describirse como un código de verificación de paridad). Para reenunciar la propiedad 2 en términos matemáticos, dejamos que la n-tupla $(c_0, c_1, \ldots, c_{n-1})$ denote una palabra de código de un código de bloques lineal (n, k). El código es cíclico si las n-tuplas

$$\begin{aligned}
&(c_{n-1}, c_0, \ldots, c_{n-2}),\\
&(c_{n-2}, c_{n-1}, \ldots, c_{n-3}),\\
&\vdots\\
&(c_1, c_2, \ldots, c_{n-1}, c_0)
\end{aligned}$$

son todas palabras de código en el código.

Para desarrollar las propiedades algebraicas de los códigos cíclicos, utilizamos los elementos $c_0, c_1, \ldots, c_{n-1}$ de una palabra de código para definir el *polinomio del código*

$$c(X) = c_0 + c_1 X + c_2 X^2 + \cdots + c_{n-1} X^{n-1} \tag{10.27}$$

donde X es una indeterminación. Naturalmente, para los códigos binarios, los coeficientes son unos y ceros. Cada potencia de X en el polinomio $c(X)$ representa un *corrimiento* a tiempo de un bit. Por tanto, la multiplicación del polinomio $c(X)$ por X puede verse como un corrimiento hacia la derecha. La pregunta clave es: ¿cómo hacer *cíclico* un corrimiento de estas características? La respuesta a esta cuestión se describe en seguida.

Considere que el polinomio de código $c(X)$ se multiplica por X^i, produciendo

$$\begin{aligned}
X^i c(X) &= X^i (c_0 + c_1 X + \cdots + c_{n-i-1} X^{n-i-1} + c_{n-i} X^{n-i}\\
&\quad + \cdots + c_{n-1} X^{n-1})\\
&= c_0 X^i + c_1 X^{i+1} + \cdots + c_{n-i-1} X^{n-1} + c_{n-i} X^n\\
&\quad + \cdots + c_{n-1} X^{n+i-1}\\
&= c_{n-i} X^n + \cdots + c_{n-1} X^{n+i-1} + c_0 X^i + c_1 X^{i+1}\\
&\quad + \cdots + c_{n-i-1} X^{n-1}
\end{aligned} \tag{10.28}$$

donde, en la última línea, hemos reordenado términos. Admitiendo, por ejemplo, que $c_{n-i} + c_{n-i} = 0$ en la suma módulo 2, es posible manipular los primeros i términos de la ecuación (10.28) del modo siguiente:

$$\begin{aligned}
X^i c(X) &= c_{n-i} + \cdots + c_{n-1} X^{i-1} + c_0 X^i + c_1 X^{i+1} + \cdots + c_{n-i-1} X^{n-1}\\
&\quad + c_{n-i}(X^n + 1) + \cdots + c_{n-1} X^{i-1}(X^n + 1)
\end{aligned} \tag{10.29}$$

Después de esto, introducimos las siguientes definiciones:

$$\begin{aligned}
c^{(i)}(X) &= c_{n-i} + \cdots + c_{n-1} X^{i-1} + c_0 X^i + c_1 X^{i+1}\\
&\quad + \cdots + c_{n-i-1} X^{n-1}
\end{aligned} \tag{10.30}$$

$$q(X) = c_{n-i} + c_{n-i+1} X + \cdots + c_{n-1} X^{i-1} \tag{10.31}$$

Por tanto, la ecuación (10.29) se reformula en la forma compacta

$$X^i c(X) = q(X)(X^n + 1) + c^{(i)}(X) \tag{10.32}$$

El polinomio $c^{(i)}(X)$ se reconoce como el polinomio de código de la palabra de código (c_{n-i}, ..., c_{n-1}, c_0, c_1, ..., c_{n-i-1}) obtenida al aplicar i corrimientos cíclicos a la palabra de código (c_0, c_1, ..., c_{n-i-1}, c_{n-i}, ..., c_{n-1}). Además, a partir de la ecuación (10.32) observamos de inmediato que $c^{(i)}(X)$ es el residuo que resulta al dividir $X^i c(X)$ entre $(X^n + 1)$. De ese modo es posible establecer formalmente la propiedad cíclica en notación de polinomios de la forma siguiente: *si $c(X)$ es un polinomio de código, entonces el polinomio*

$$c^{(i)}(X) = X^i c(X) \bmod (X^n + 1) \tag{10.33}$$

también es un polinomio de código para todo corrimiento cíclico i; el término *mod* es la abreviación de *módulo*. La forma especial de la multiplicación de polinomio descrita en la ecuación (10.33) se conoce como *módulo de multiplicación* $X^n + 1$. En efecto, la multiplicación está sujeta a la restricción $X^n = 1$, cuya aplicación regresa el polinomio $X^i c(X)$ al orden $n - 1$ para toda $i < n$. (Advierta que en la aritmética de módulo 2, $X^n + 1$ tiene el mismo valor que $X^n - 1$.)

■ Polinomio generador

El polinomio $X^n + 1$ y sus factores desempeñan un papel fundamental en la generación de códigos cíclicos. Sea $g(X)$ un polinomio de grado $n - k$ que es un factor de $X^n + 1$; como tal, $g(X)$ es *el polinomio de menor grado en el código*. En general, $g(X)$ puede expandirse del modo siguiente:

$$g(X) = 1 + \sum_{i=1}^{n-k-1} g_i X^i + X^{n-k} \tag{10.34}$$

donde el coeficiente g_i es igual a 0 o 1. De acuerdo con esta expansión, el polinomio $g(X)$ tiene dos términos con coeficiente 1 separados por $n - k - 1$ términos. El polinomio $g(X)$ recibe el nombre de *polinomio generador* de un código cíclico. Este último está determinado en forma única por el polinomio generador $g(X)$ en cuanto a que cada polinomio de código en el código puede expresarse en la forma de un producto polinomial de la siguiente manera:

$$c(X) = a(X)g(X) \tag{10.35}$$

donde $a(X)$ es un polinomio en X con grado $k - 1$. Las $c(X)$ formadas de ese modo satisfacen la condición de la ecuación (10.33), pues $g(X)$ es un factor de $X^n + 1$.

Supóngase que se nos da el polinomio generador $g(X)$ y que el requerimiento es codificar la secuencia de mensaje (m_0, m_1, ..., m_{k-1}) en un código cíclico *sistemático* (n, k). Es decir, los bits de mensaje se transmiten en forma inalterada, como indica la siguiente estructura para una palabra de código (vea la figura 10.4):

$$(\underbrace{b_0, b_1, \ldots, b_{n-k-1}}_{n-k \text{ bits de paridad}}, \quad \underbrace{m_0, m_1, \ldots, m_{k-1}}_{k \text{ bits de mensaje}})$$

Defínase el *polinomio de mensaje* por medio de

$$m(X) = m_0 + m_1 X + \ldots + m_{k-1} X^{k-1} \tag{10.36}$$

y considérese que

$$b(X) = b_0 + b_1 X + \ldots + b_{n-k-1} X^{n-k-1} \tag{10.37}$$

De acuerdo con la ecuación (10.1), deseamos que el polinomio de código esté en la forma

$$c(X) = b(X) + X^{n-k}m(X) \tag{10.38}$$

Por tanto, el empleo de las ecuaciones (10.35) y (10.38) produce

$$a(X)g(X) = b(X) + X^{n-k}m(X)$$

De manera equivalente, a la luz de la adición módulo 2, es posible escribir

$$\frac{X^{n-k}m(X)}{g(X)} = a(X) + \frac{b(X)}{g(X)} \tag{10.39}$$

La ecuación (10.39) establece que el polinomio $b(X)$ es el *residuo* que resulta después de dividir $X^{n-k}m(X)$ entre $g(X)$.

Podemos resumir ahora los pasos implicados en el procedimiento de codificación correspondiente a un código cíclico (n, k) asegurado de una estructura sistemática. De manera específica, procedemos de la siguiente forma:

1. Se multiplica el polinomio de mensaje $m(X)$ por X^{n-k}.
2. Se divide $X^{n-k}m(X)$ entre el polinomio generador $g(X)$, obteniéndose el residuo $b(X)$.
3. Se suma $b(X)$ a $X^{n-k}m(X)$, obteniéndose el polinomio de código $c(X)$.

■ POLINOMIO DE VERIFICACIÓN DE PARIDAD

Un código cíclico (n, k) está especificado en forma única por su polinomio generador $g(X)$ de orden $(n - k)$. Un código de este tipo se especifica también en forma única mediante otro polinomio de grado k, que se denomina *polinomio de verificación de paridad*, definido por

$$h(X) = 1 + \sum_{i=1}^{k-1} h_i X^i + X^k \tag{10.40}$$

donde los coeficientes h_i son 0 o 1. El polinomio de verificación de paridad $h(X)$ tiene una forma similar al polinomio generador en cuanto a que hay dos términos con coeficiente 1, pero separados por $k - 1$ términos.

El polinomio generador $g(X)$ es equivalente a la matriz generadora $\mathbf{G}$ como una descripción del código. En forma correspondiente, el polinomio de verificación de paridad, denotado por $h(X)$, es una representación equivalente de la matriz de verificación de paridad $\mathbf{H}$. De ese modo encontramos que la relación de matriz $\mathbf{HG}^T = \mathbf{0}$ presentada en la ecuación (10.15) para códigos de bloque lineales corresponde a la relación

$$g(X)h(X) \bmod (X^n + 1) = 0 \tag{10.41}$$

Por tanto, podemos establecer que *el polinomio generador* $g(X)$ *y el polinomio de verificación de paridad* $h(X)$ *son factores del polinomio* $X^n + 1$, como indica

$$g(X)h(X) = X^n + 1 \tag{10.42}$$

Esta propiedad proporciona la base para elegir el polinomio generador o de verificación de paridad de un código cíclico. En particular, podemos establecer que si $g(X)$ es un polinomio de grado $(n - k)$ y es también un factor de $X^n + 1$, entonces $g(X)$ es el polinomio generador de un código cíclico (n, k). De manera equivalente, es factible establecer que si $h(X)$ es un polinomio de grado k y es también un factor de $X^n + 1$, entonces $h(X)$ es el polinomio de verificación de paridad de un código cíclico (n, k).

Cabe un comentario final. Todo factor de $X^n + 1$ con grado $(n - k)$, el número de bits de paridad, puede utilizarse como un polinomio generador. Para valores grandes de n, es posible que el polinomio $X^n + 1$ tenga muchos factores de grado $n - k$. Algunos de estos factores

polinomiales generan buenos códigos cíclicos, en tanto que algunos generan códigos cíclicos inadecuados. La cuestión de cómo elegir polinomios generadores que produzcan buenos códigos cíclicos es muy difícil de resolver. En realidad, los teóricos de la codificación han dedicado un gran esfuerzo en la búsqueda de buenos códigos cíclicos.

Matrices generadoras y de verificación de paridad

Dado el polinomio generador $g(X)$ de un código cíclico (n, k), podemos construir la matriz generadora **G** del código advirtiendo que los k polinomios $g(X)$, $Xg(X)$, ..., $X^{k-1}g(X)$ expanden el código. Por tanto, las n–tuplas correspondientes a estos polinomios pueden utilizarse como renglones de la matriz generadora **G** de k por n.

Sin embargo, la construcción de la matriz de verificación de paridad **H** del código cíclico a partir del polinomio de verificación de paridad $h(X)$ requiere especial atención, como se describe aquí. Al multiplicar la ecuación (10.42) por $a(x)$ y al utilizar después la ecuación (10.35), obtenemos

$$c(X)h(X) = a(X) + X^n a(X) \tag{10.43}$$

Los polinomios $c(X)$ y $h(X)$ son a su vez definidos por las ecuaciones (10.27) y (10.40), respectivamente, lo cual significa que su producto en el primer miembro de la ecuación (10.43) contiene términos con potencias que se extienden hasta $n + k - 1$. Por otra parte, el polinomio $a(X)$ tiene grado $k - 1$ o menor, lo que implica que las potencias X^k, X^{k+1}, ..., X^{n+1} *no* aparecen en el polinomio en el segundo miembro de la ecuación (10.43). De ese modo, igualando a cero los coeficientes de X^k, X^{k-1}, ..., X^{n-1} en la expansión del polinomio del producto $c(X)h(X)$, obtenemos el siguiente conjunto de $n - k$ ecuaciones:

$$\sum_{i=j}^{j+k} c_i h_{k+j-i} = 0 \quad \text{para } 0 \le j \le n-k-1 \tag{10.44}$$

Comparando la ecuación (10.44) con la relación correspondiente de la ecuación (10.16), podemos señalar la siguiente observación importante: los coeficientes del polinomio de verificación de paridad $h(X)$ incluidos en la multiplicación de polinomios descrita en la ecuación (10.44) se arreglan en *orden invertido* con respecto a los coeficientes de la matriz de verificación de paridad **H** implicada en la formación del producto interior de los vectores que se describe en la ecuación (10.16). Esta observación sugiere que definamos el *recíproco del polinomio de verificación de paridad* del modo siguiente:

$$\begin{aligned} X^k h(X^{-1}) &= X^k\left(1 + \sum_{i=1}^{k-1} h_i X^{-i} + X^{-k}\right) \\ &= 1 + \sum_{i=1}^{k-1} h_{k-i} X^i + X^k \end{aligned} \tag{10.45}$$

que también es un factor de $X^n + 1$. Las n-tuplas pertenecientes a los polinomios $(n - k)\ X^k h(X^{-1})$, $X^{k+1}h(X^{-1})$, ..., $X^{n-1}h(X^{-1})$ pueden utilizarse ahora en los renglones de la matriz de verificación de paridad **H** de $(n - k)$ por n.

En general, la matriz generadora **G** y la matriz de verificación de paridad **H** construidas de la manera descrita aquí no están en sus formas sistemáticas. Pueden ponerse en esas formas realizando operaciones simples sobre sus renglones respectivos, como se ilustra en el ejemplo 10.3.

Codificador para códigos cíclicos

Demostramos antes que el procedimiento de codificación para un código cíclico (n, k) en una forma sistemática incluye tres pasos: 1) multiplicación de los polinomios de mensaje $m(X)$ por

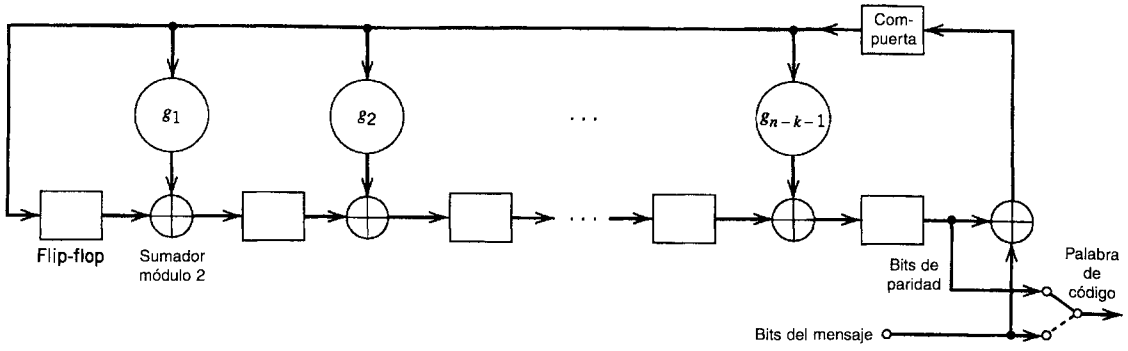


FIGURA 10.8 Codificador de un modo cíclico (n, k).

X^{n-k}, 2) división de $X^{n-k}m(X)$ por el polinomio generador $g(X)$ para obtener el residuo $b(X)$ y 3) suma de $b(X)$ a $X^{n-k}m(X)$ para formar el polinomio de código deseado. Estos tres pasos pueden ejecutarse por medio del codificador que se presenta en la figura 10.8, consistente en un *registro de corrimiento realimentado lineal* con $(n - k)$ etapas.

Las cajas en la figura 10.8 representan *flip-flops* o *elementos de retardo unitario*. El flip-flop es un dispositivo que reside en uno de dos posibles estados denotados por 0 y 1. Un *reloj externo* (no indicado en la figura 10.8) controla la operación de todos los flip-flops. Cada vez que el reloj marca, el contenido de los flip-flops (inicialmente establecido en el estado 0) se corren hacia la dirección de las flechas. Además de los flip-flops, el codificador de la figura 10.8 incluye un segundo conjunto de elementos lógicos, es decir, *sumadores*, los cuales calculan las sumas módulo 2 de sus respectivas entradas. Por último, los *multiplicadores* multiplican sus respectivas entradas por los coeficientes asociados. En particular, si el coeficiente $g_i = 1$, el multiplicador se encuentra justamente en una "conexión" directa. Si, en cambio, el coeficiente $g_i = 0$, el multiplicador está en "sin conexión".

La operación del codificador que se muestra en la figura 10.8 procede de la manera siguiente:

1. La compuerta se activa. Por tanto, los k bits de mensaje se corren hacia el canal. Tan rápido como los k bits de mensaje han entrado al registro de corrimiento, los $(n - k)$ bits resultantes en el registro forman los bits de paridad [recuerde que los bits de paridad son los mismos que los coeficientes del residuo $b(X)$].
2. La compuerta se desactiva, interrumpiendo las conexiones de realimentación.
3. Los contenidos del registro de corrimiento se leen en el canal.

■ CÁLCULO DEL SÍNDROME

Suponga que la palabra de código $(c_0, c_1, ..., c_{n-1})$ se transmite por un canal ruidoso, originando una palabra recibida $(r_0, r_1, ..., r_{n-1})$. De la sección 10.3 recordamos que el primer paso en la decodificación de un código de bloques lineal consiste en calcular el síndrome de la palabra recibida. Si el síndrome es cero, no hay errores de transmisión en la palabra recibida. Si, en cambio, el síndrome es distinto de cero, la palabra recibida contiene errores de transmisión que es necesario corregir.

En el caso de un código cíclico en la forma sistemática, el síndrome puede calcularse con facilidad. Considere que la palabra recibida se representa por medio de un polinomio de grado $n - 1$ o menor, como indica

$$r(X) = r_0 + r_1X + \ldots + r_{n-1}X^{n-1} \tag{10.46}$$

Sea $q(X)$ el cociente y $s(X)$ el residuo, los cuales resultan de dividir $r(X)$ entre el polinomio generador $g(X)$. Por tanto, es posible expresar $r(X)$ en la forma siguiente:

$$r(X) = q(X)g(X) + s(X) \tag{10.47}$$

El residuo $s(X)$ es un polinomio de grado $n - k - 1$ o menor, que es el resultado de interés. Se le denomina *polinomio del síndrome* debido a que sus coeficientes integran el síndrome **s** de $(n - k)$ por 1.

La figura 10.9 muestra un *calculador de síndrome* que es idéntico al codificador de la figura 10.8, salvo por el hecho de que los bits recibidos se alimentan a las $(n - k)$ etapas del registro de corrimiento realimentado desde la izquierda. Tan pronto como los bits recibidos se han corrido hacia el registro de corrimiento, sus contenidos definen al síndrome **s**.

El polinomio del síndrome $s(X)$ tiene las siguientes propiedades útiles que surgen a partir de la definición dada en la ecuación (10.47).

1. *El síndrome de un polinomio de palabra recibida es también el síndrome del correspondiente polinomio de error.*

Dado que un código cíclico con polinomio $c(X)$ se envía por un canal ruidoso, el polinomio de la palabra recibida se define mediante

$$r(X) = c(X) + e(X) \tag{10.48}$$

donde $e(X)$ es el *polinomio de error*. De modo equivalente, podemos escribir

$$e(X) = r(X) + c(X) \tag{10.49}$$

Por tanto, al sustituir las ecuaciones (10.35) y (10.47) en la (10.49), obtenemos

$$e(X) = u(X)g(X) + s(X) \tag{10.50}$$

donde el cociente es $u(X) = a(X) + q(X)$. La ecuación (10.50) muestra que $s(X)$ es también el síndrome del polinomio de error $e(X)$. La implicación de esta propiedad es que cuando el polinomio del síndrome $s(X)$ es distinto de cero, se detecta la presencia de errores de transmisión en la palabra recibida.

2. *Sea $s(X)$ el síndrome de un polinomio de palabra recibida $r(X)$. Entonces, el síndrome de $Xr(X)$, un corrimiento cíclico de $r(X)$, es $Xs(X)$.*

Al aplicar el corrimiento cíclico en ambos miembros de la ecuación (10.47) obtenemos

$$Xr(X) = Xq(X)g(X) + Xs(X) \tag{10.51}$$

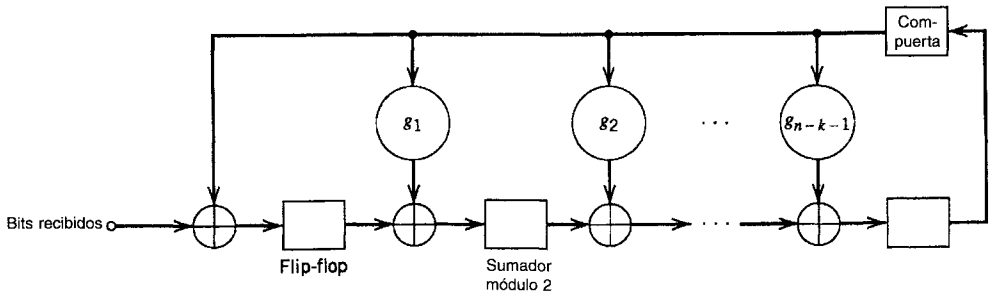


FIGURA 10.9 Calculador del síndrome para código cíclico (n, k).

a partir de lo cual vemos de inmediato que $Xs(X)$ es el residuo de la división de $Xr(X)$ entre $g(X)$. Por tanto, el síndrome de $Xr(X)$ es $Xs(X)$, como se había establecido. Es posible generalizar este resultado señalando que si $s(X)$ es el síndrome de $r(X)$, entonces $X^i s(X)$ es el síndrome de $X^i r(X)$.

3. *El polinomio del síndrome $s(X)$ es idéntico al polinomio de error $e(X)$, suponiendo que los errores están confinados a los $(n-k)$ bits de verificación de paridad del polinomio de la palabra recibida $r(X)$.*

La suposición hecha aquí es otra forma de afirmar que el grado del polinomio de error $e(X)$ es menor o igual que $(n-k-1)$. Puesto que el polinomio generador $g(X)$ es de grado$(n-k)$, por definición, se concluye que la ecuación (10.50) sólo puede ser satisfecha si el cociente $u(X)$ es cero. En otras palabras, el polinomio de error $e(X)$ y el polinomio del síndrome $s(X)$ son el mismo. La implicación de la propiedad 3 es que, bajo las condiciones mencionadas antes, la corrección del error puede conseguirse sumando simplemente el polinomio del síndrome $s(X)$ al polinomio de la palabra recibida $r(X)$.

EJEMPLO 10.3 Reexposición de los códigos de Hamming

Para ilustrar los aspectos relacionados con la representación en polinomios de los códigos cíclicos, consideramos la generación de un código cíclico (7, 4). Con la longitud de bloque $n = 7$, iniciando a partir de la factorización de $X^7 + 1$ en tres *polinomios irreducibles*:

$$X^7 + 1 = (1 + X)(1 + X^2 + X^3)(1 + X + X^3)$$

Por un "polinomio irreducible" queremos decir un polinomio que no puede factorizarse utilizando únicamente polinomios con coeficientes que provienen del campo binario. Un polinomio irreducible de grado m se dice que será *primitivo* si el entero positivo n más pequeño para el cual el polinomio divide a $X^n + 1$ es $n = 2^m - 1$. Para el ejemplo presente, los dos polinomios $(1 + X^2 + X^3)$ y $(1 + X + X^3)$ son primitivos. Consideremos

$$g(X) = 1 + X + X^3$$

como el polinomio generador, cuyo grado es igual al número de bits de paridad. Esto equivale a que el polinomio de verificación de paridad esté dado por

$$\begin{aligned} h(X) &= (1 + X)(1 + X^2 + X^3) \\ &= 1 + X + X^2 + X^4 \end{aligned}$$

cuyo grado es igual al número de bits de mensaje $k = 4$.

Después de esto, ilustramos el procedimiento para la construcción de una palabra de código utilizando este polinomio generador para codificar la secuencia de mensaje 1001. El polinomio de mensaje correspondiente está definido por

$$m(X) = 1 + X^3$$

De ese modo, al multiplicar $m(X)$ por $X^{n-k} = X^3$, se llega a

$$X^{n-k} m(X) = X^3 + X^6$$

El segundo paso es dividir $X^{n-k}m(X)$ entre $g(X)$, los detalles de lo cual (para el ejemplo presente) se indican en seguida:

$$\begin{array}{r} X^3 + X \\ X^3 + X + 1 \overline{)\, X^6 \qquad + X^3} \\ \underline{X^6 \quad + X^4 + X^3} \\ X^4 \\ \underline{X^4 \quad + X^2 + X} \\ X^2 + X \end{array}$$

Advierta que en esta división larga hemos tratado la sustracción del mismo modo que la adición, puesto que estamos operando en aritmética de módulo 2. En ese caso podemos escribir

$$\frac{X^3 + X^6}{1 + X + X^3} = X + X^3 + \frac{X + X^2}{1 + X + X^3}$$

Es decir, el cociente $a(X)$ y el residuo $b(X)$ son los siguientes, respectivamente:

$$a(X) = X + X^3$$
$$b(X) = X + X^2$$

Por tanto, de acuerdo con la ecuación (10.38) encontramos que el polinomio de código que se desea es

$$c(X) = b(X) + X^{n-k}m(X)$$
$$= X + X^2 + X^3 + X^6$$

La palabra de código es entonces 0111001. Los cuatro bits más a la derecha, 1001, son los bits de mensaje especificados. Los tres bits más a la izquierda, 011, corresponden a los bits de verificación de paridad. La palabra de código generada de este modo es exactamente la misma que la correspondiente que se muestra en la tabla 10.1 para un código de Hamming (7, 4).

Es posible generalizar este resultado afirmando que *todo código cíclico generado por un polinomio primitivo es un código de Hamming de distancia mínima 3.*

En seguida demostramos que el polinomio generador $g(X)$ y el polinomio de verificación de paridad $h(X)$ especifican en forma única a la matriz generadora **G** y a la matriz de verificación de paridad **H**, respectivamente.

Para construir la matriz generadora **G** de 4 por 7, empezamos con cuatro polinomios representados por $g(X)$ y tres versiones de corrimiento cíclico de ella, según se indica mediante

$$g(X) = 1 + X + X^3$$
$$Xg(X) = X + X^2 + X^4$$
$$X^2g(X) = X^2 + X^3 + X^5$$
$$X^3g(X) = X^3 + X^4 + X^6$$

Los polinomios $g(X)$, $Xg(X)$, $X^2g(X)$ y $X^3g(X)$ representan polinomios de código en el código de Hamming (7, 4). Si los coeficientes de estos polinomios se utilizan como los elementos de los renglones de una matriz de 4 por 7, obtenemos la siguiente matriz generadora:

$$\mathbf{G}' = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Claramente, la matriz generadora **G′** construida de ese modo no está en forma sistemática. Podemos ponerla en esa forma sumando el primer renglón al tercero y agregando la suma de los primeros dos renglones al renglón cuarto. Estas manipulaciones producen la matriz generadora que se desea:

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

que es exactamente la misma que la del ejemplo 10.2.

A continuación mostramos cómo construir la matriz de verificación de paridad **H** de 3 por 7 a partir del polinomio de verificación de paridad $h(X)$. Para efectuar lo anterior, tomamos primero el *recíproco* de $h(X)$, a saber, $X^4h(X^{-1})$. Para el problema presente, formamos tres polinomios representados por $X^4h(X^{-1})$ y dos versiones recorridas de ésta, en la forma que indica

$$X^4h(X^{-1}) = 1 + X^2 + X^3 + X^4$$
$$X^5h(X^{-1}) = X + X^3 + X^4 + X^5$$
$$X^6h(X^{-1}) = X^2 + X^4 + X^5 + X^6$$

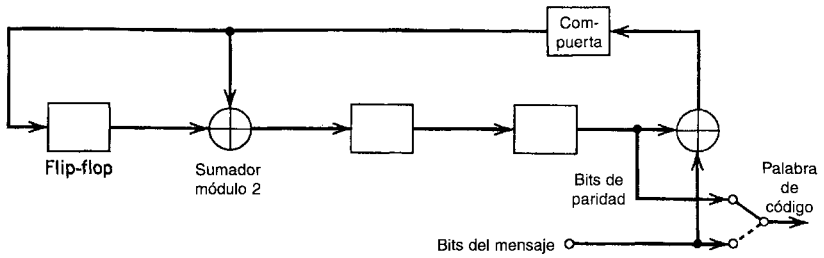


FIGURA 10.10 Codificador para el código cíclico (7, 4) generado por $g(X) = 1 + X + X^3$.

Utilizando los coeficientes de estos tres polinomios como los elementos de los renglones de la matriz de verificación de paridad de 3 por 7, obtenemos

$$\mathbf{H}' = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

En este caso también vemos que la matriz $\mathbf{H}'$ no está en forma sistemática. Para que lo esté, sumamos el tercer renglón al primero para llegar a

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

que es exactamente la misma que la del ejemplo 10.2.

La figura 10.10 muestra el codificador para el código de Hamming cíclico (7, 4) generado por un polinomio $g(X) = 1 + X + X^3$. Para ilustrar la operación de este codificador, considere la secuencia de mensaje (1001). Los contenidos del registro de corrimiento se modifican mediante los bits de mensaje entrantes como en la tabla 10.3. Después de cuatro corrimientos, los contenidos del registro de corrimiento y consecuentemente los bits de paridad, son (011). Por tanto, agregando estos bits de paridad a los bits de mensaje (1001), obtenemos la palabra de código (0111001); este resultado es exactamente el mismo que se determinó antes en el ejemplo.

La figura 10.11 presenta el correspondiente calculador del síndrome para el código de Hamming (7, 4). Considere que la palabra de código transmitida es (0111001), y la recibida (0110001); es decir, el bit de en medio es erróneo. Como los bits recibidos se alimentan en el registro de corrimiento, inicialmente fijado en cero, sus contenidos se modifican como en la tabla 10.4. Al final del séptimo corrimiento, el síndrome se identifica a partir de los contenidos del registro de corrimiento igual a 110. Puesto que el síndrome es distinto de cero, la palabra recibida es errónea. Además, de acuerdo con la tabla 10.2, vemos que un patrón de errores correspondientes a este síndrome es 0001000. Lo anterior indica que el error está en el bit de en medio de la palabra recibida, lo cual es realmente el caso. ◀

TABLA 10.3 ***Contenidos del registro de corrimiento en el codificador de la figura 10.10 para la secuencia del mensaje (1001)***

Corrimiento	*Bit de entrada*	*Contenidos del registro*
		0 0 0 (estado inicial)
1	1	1 1 0
2	0	0 1 1
3	0	1 1 1
4	1	0 1 1

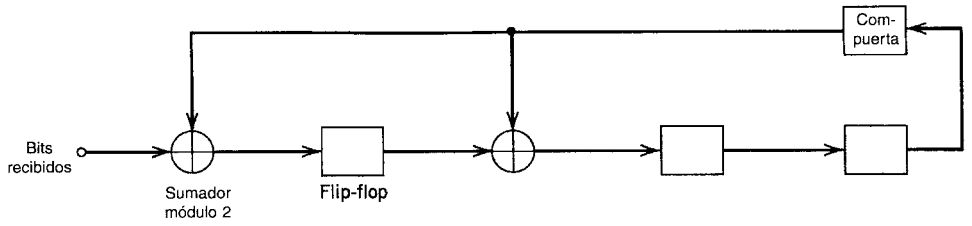


FIGURA 10.11 Calculador del síndrome para el código cíclico (7, 4) generado por el polinomio $g(X) = 1 + X + X^3$.

EJEMPLO 10.4 Códigos de longitud máxima

Para cualquier entero positivo $m \geq 3$, existe un *código de longitud máxima* con los siguientes parámetros:

Longitud de bloque:	$n = 2^m - 1$
Número de bits del mensaje:	$k = m$
Número de bits de paridad:	$d_{\text{mín}} = 2^{m-1}$

Los códigos de longitud máxima se generan mediante polinomios de la forma

$$g(X) = \frac{1 + X^n}{h(X)} \tag{10.52}$$

donde $h(X)$ es un polinomio primitivo de grado m. Antes señalamos que cualquier código cíclico generado por un polinomio primitivo es un código de Hamming de distancia mínima 3 (vea el ejemplo 10.3). Por tanto, se concluye que los códigos de longitud máxima son los *duales* de los códigos de Hamming.

El polinomio $h(X)$ define las conexiones de realimentación del codificador. El polinomio generador $g(X)$ define un periodo del código de longitud máxima, suponiendo que el codificador está en el estado inicial 00 ... 01. Para ilustrar estos puntos, considere el ejemplo de un código de longitud máxima (7, 3), que es el dual del código de Hamming (7, 4) que se describe en el ejemplo 10.3. De ese modo, al elegir

$$h(X) = 1 + X + X^3$$

encontramos que el polinomio generador del código de longitud máxima (7, 3) es

$$g(X) = 1 + X + X^2 + X^4$$

TABLA 10.4 ***Contenidos del calculador del síndrome en la figura 10.11 para la palabra recibida 0110001***

Corrimiento	*Bit de entrada*	*Contenidos del registro de corrimiento*
		0 0 0 (estado inicial)
1	1	1 0 0
2	0	0 1 0
3	0	0 0 1
4	0	1 1 0
5	1	1 1 1
6	1	0 0 1
7	0	1 1 0

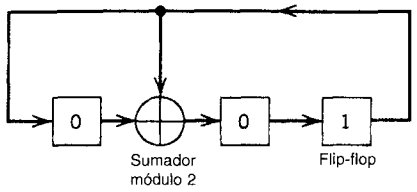


FIGURA 10.12 Codificador para el código de longitud máxima (7, 3); el estado inicial del codificador se muestra en la figura.

La figura 10.12 presenta el codificador para el código de longitud máxima (7, 3), cuyas conexiones de realimentación son exactamente las mismas que se presentan en la figura 8.2. El periodo del código es $n = 7$. Por tanto, suponiendo que el codificador se encuentre en el estado inicial 001, como se indica en la figura 10.12, encontramos que la secuencia de salida está descrita por

$$\underbrace{1\ 0\ 0}_{\text{estado inicial}} \qquad \underbrace{1\ 1\ 1\ 0\ 1\ 0\ 0}_{g(X)=1+X+X^2+X^4}$$

Este resultado puede validarse sin dificultad efectuando un círculo en torno al codificador de la figura 10.12.

Advierta que si hubiéramos elegido el otro polinomio primitivo

$$h(X) = 1 + X^2 + X^3$$

para el código de longitud máxima (7, 3), obtendríamos simplemente la "imagen" del código descrito antes, y la secuencia de salida estaría "invertida" en el tiempo. ◀

■ OTROS CÓDIGOS CÍCLICOS

Concluimos la exposición de los códigos cíclicos presentando las características de otras tres clases importantes de códigos cíclicos.

Códigos cíclicos de verificación de redundancia

Los códigos cíclicos son sumamente adecuados para la *detección de errores*. Hacemos este planteamiento por dos razones. En primer lugar, pueden diseñarse para detectar muchas combinaciones de errores probables. En segundo, la ejecución de los circuitos tanto de codificación como de detección de errores es práctica. Por estas razones muchos códigos de detección de errores que se utilizan en la práctica son de tipo cíclico. Un código cíclico empleado para la detección de errores se conoce como *código cíclico de verificación* de *redundancia* (CRC).

Definimos una *ráfaga de errores* de longitud B en una palabra recibida de n bits como una secuencia contigua de B bits en la cual el primero y el último bits o cualquier número de bits intermedios se reciben con errores. Los códigos CRC binarios (n, k) son capaces de detectar los siguientes patrones de error:

1. Todas las ráfagas de errores de longitud $n - k$ o inferior.
2. Una fracción de ráfagas de errores de longitud igual a $n - k + 1$; la fracción es igual a $1 - 2^{-(n-k-1)}$.
3. Una fracción de ráfagas de errores de longitud mayor que $n - k + 1$; la fracción es igual a $1 - 2^{-(n-k-1)}$.
4. Todas las combinaciones de $d_{\text{mín}} - 1$ (o menos) errores.
5. Todos los patrones de errores con número impar de errores si el polinomio generador $g(X)$ para el código tiene un número par de coeficientes distintos de cero.

TABLA 10.5 *Códigos CRC*

Código	*Polinomio generador, g(X)*	$n-k$
Código CRC-12	$1 + X + X^2 + X^3 + X^{11} + X^{12}$	12
Código CRC-16 (EU)	$1 + X^2 + X^{15} + X^{16}$	16
Código CRC-ITU	$1 + X^5 + X^{12} + X^{16}$	16

La tabla 10.5 presenta los polinomios generadores de tres códigos CRC que se han convertido en estándares internacionales. Los tres códigos contienen $1 + X$ como un factor principal. El código CRC-12 se utiliza para caracteres de 6 bits y los otros dos códigos se emplean para caracteres de 8 bits. Los códigos CRC proporcionan un poderoso método de detección de errores aplicable en estrategias de solicitud de repetición automática (ARQ) explicadas en la sección 10.1, así como en las líneas de suscriptor digital que se analizaron en el capítulo 4.

Códigos Bose-Chaudhuri-Hocquenghem (BCH)[5]

Los *códigos BCH* constituyen una de las clases más importantes y poderosas de códigos de bloques lineales, los cuales presentan una amplia variedad de parámetros. Los códigos BCH binarios más comunes, conocidos como *códigos BCH primitivos*, están caracterizados por cualesquiera enteros positivos m (igual o mayor que 3) y t [menor que $(2^m - 1)/2$] mediante los siguientes parámetros:

Longitud de bloque: $n = 2^m - 1$
Número de bits del mensaje: $k \geq n - mt$
Distancia mínima: $d_{\text{mín}} \geq 2t + 1$

Cada código BCH es un *código de corrección de t errores* en el sentido de que puede detectar y corregir hasta t errores aleatorios por palabra de código. Es posible describir los códigos de corrección de un solo error de Hamming como códigos BCH. Estos últimos ofrecen flexibilidad en la elección de los parámetros de código. A saber, longitud de bloque y tasa de código. Además, en longitudes de bloque de unos cuantos cientos de bits o menos, los códigos BCH se encuentran entre los más conocidos de la misma longitud de bloque y tasa de código.

Un tratamiento detallado de la construcción de los códigos BCH está más allá del objetivo de nuestra presente explicación. Para proporcionar una idea de su capacidad, en la tabla 10.6 presentamos los parámetros de código y los polinomios generadores correspondientes a códigos BCH de bloque binarios de longitud hasta $2^5 - 1$. Por ejemplo, suponga que deseamos

TABLA 10.6 *Códigos BCH binarios de longitud hasta de $2^5 - 1$*

n	k	t	*Polinomio generador*								
7	4	1								1	011
15	11	1								10	011
15	7	2							111	010	001
15	5	3						10	100	110	111
31	26	1								100	101
31	21	2						11	101	101	001
31	16	3					1	000	111	110	101	111
31	11	5				101	100	010	011	011	010	101
31	6	7	11	001	011	011	110	101	000	100	111

Notación: n = longitud del bloque
k = número de bits del mensaje
t = número máximo de errores detectables
Los coeficientes de orden superior del polinomio generador $g(X)$ están a la izquierda.

construir los polinomios generadores para el código BCH (15, 7). De acuerdo con la tabla 10.6 tenemos (111 010 001) para los coeficientes del polinomio generador; de ese modo, escribimos

$$g(X) = X^8 + X^7 + X^6 + X^4 + 1$$

Códigos Reed-Solomon[6]

Los códigos *Reed-Solomon* constituyen una subclase importante de códigos BCH *no binarios*; se abrevian en ocasiones como códigos RS. El codificador para un código RS difiere de un codificador binario en el sentido de que éste opera sobre bits múltiples en vez de bits individuales. Específicamente, un código RS (n, k) se utiliza para codificar símbolos de m bits en bloques consistentes de $n = 2^m - 1$ símbolos, es decir, $m(2^m - 1)$ bits, donde $m \geq 1$. Por tanto, el algoritmo de codificación expande un bloque de k símbolos a n símbolos agregando $n - k$ símbolos redundantes. Cuando m es una potencia entera de dos, los símbolos de m bits reciben el nombre de *bytes*. Un valor popular de m es 8; en realidad, los códigos RS de 8 bits son extremadamente poderosos.

Un código RS de corrección de t errores tiene los siguientes parámetros:

Longitud de bloque:	$n = 2^m - 1$ símbolos
Tamaño del mensaje:	k símbolos
Tamaño de la verificación de paridad:	$n - k = 2t$ símbolos
Distancia mínima:	$d_{\text{mín}} = 2t + 1$ símbolos

La longitud de bloque del código RS es una unidad menor que el tamaño de un símbolo de código, y la distancia mínima es mayor en uno que el número de símbolos de verificación de paridad. Los códigos RS usan con gran eficiencia la redundancia, y las longitudes de bloque y los tamaños de símbolo pueden ajustarse fácilmente para alojar una amplia gama de tamaños de mensaje. Además, los códigos RS ofrecen un extenso intervalo de tasa de código que es posible elegir para optimizar el desempeño. Por último, están disponibles técnicas de decodificación eficiente que se emplean con códigos RS, lo cual es una razón más para su amplia aplicación (por ejemplo, los sistemas de audio digital de disco compacto).

10.5 *Códigos convolucionales*[7]

En la codificación de bloques, el codificador acepta un bloque de mensaje de k bits y genera una palabra de código de n bits. De ese modo, las palabras de código se producen en un esquema bloque por bloque. Claramente, deben tomarse provisiones en el codificador para retener en un búfer un bloque de mensaje completo antes de generar la palabra de código asociada. Sin embargo, hay aplicaciones en las que los bits del mensaje vienen en *forma serial* y no en bloques grandes, en cuyo caso el uso de un búfer puede ser indeseable. En este tipo de situaciones, es probable que el empleo de la *codificación convolucional* sea el método preferido. Un codificador convolucional genera bits redundantes utilizando *convoluciones módulo 2*, de ahí el nombre.

El codificador de un código convolucional binario con tasa $1/n$, medida en bits por símbolo, puede considerarse como una *máquina de estado finito* que consiste en un registro de corrimiento de M etapas con conexiones preestablecidas a n sumadores módulo 2 y a un multiplexor que pone en serie las salidas de los sumadores. Una secuencia de mensaje de L bits produce una secuencia de salida codificada de longitud $n(L + M)$ bits. La *tasa de código* está dada entonces por

$$r = \frac{L}{n(L+M)} \quad \text{bits/símbolo} \tag{10.53}$$

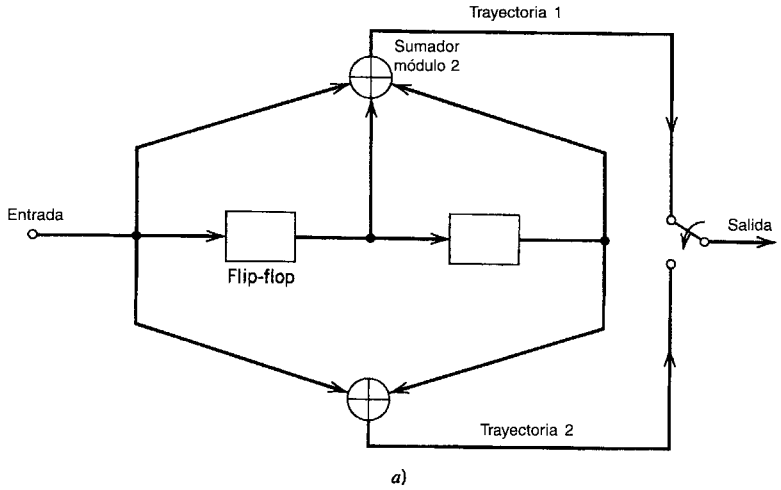
Por lo común, tenemos $L >> M$. Por tanto, la tasa de código se simplifica en

$$r \approx \frac{1}{n} \text{ bits/símbolo} \tag{10.54}$$

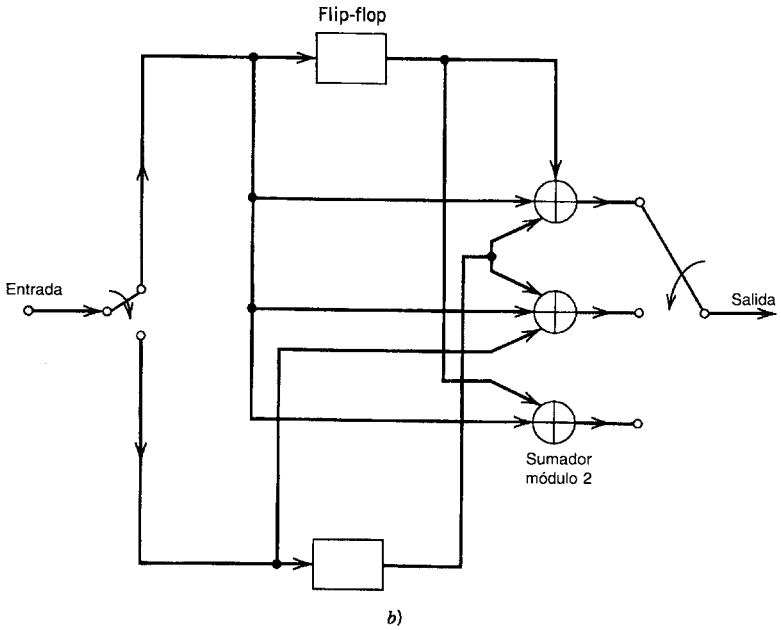
La *longitud de restricción* de un código convolucional, expresada en términos de bits de mensaje, se define como el número de corrimientos sobre el cual un bit de un solo mensaje puede influir en la salida del codificador. En un codificador con un registro de corrimiento de M etapas, la *memoria* del codificador es igual a M bits de mensaje y se requieren $K = M + 1$ corrimientos para que un bit de mensaje entre al registro de corrimiento y finalmente salga. Por tanto, la longitud de restricción del codificador es K.

La figura 10.13*a* muestra un codificador convolucional con $n = 2$ y $K = 3$. De ese modo, la tasa de código de este codificador es 1/2. El codificador de la figura 10.13*a* opera sobre la secuencia de mensaje entrante, un bit a la vez.

Es posible generar un código convolucional binario con tasa k/n utilizando k registros de corrimiento independientes con conexiones preestablecidas con n sumadores módulo 2, un



a)



b)

FIGURA 10.13 *a*) Codificador convolución de tasa 1/2 y longitud de restricción 3. *b*) Codificador convolucional de tasa 2/3 y longitud de restricción 2.

multiplexor de entrada y un multiplexor de salida. Un ejemplo de un codificador de este tipo se presenta en la figura 10.13*b*, donde $k = 2$, $n = 3$, y los dos registros de corrimiento tienen $K = 2$ cada uno. La tasa de código corresponde a 2/3. En este segundo ejemplo, el codificador procesa la secuencia de mensaje entrante dos bits a la vez.

Los códigos convolucionales generados por los codificadores de la figura 10.13 son códigos *no sistemáticos*. A diferencia de la codificación de bloques, el uso de códigos no sistemáticos se prefiere por lo regular con respecto a los códigos sistemáticos en la codificación convolucional.

Cada trayectoria que conecta la salida con la entrada de un codificador convolucional puede caracterizarse en términos de su *respuesta al impulso*, definida como la respuesta de esa trayectoria a un símbolo 1 aplicada en su entrada, con cada flip-flop en el conjunto codificador inicialmente en el estado cero. De manera equivalente, podemos caracterizar cada trayectoria en términos de un *polinomio generador*, definido como la *transformada de retorno unitario* de la respuesta al impulso. Específicamente, dejemos que la *secuencia generadora* $(g_0^{(i)}, g_1^{(i)}, g_2^{(i)}, \ldots, g_M^{(i)})$ denote la respuesta al impulso de la i-ésima trayectoria, donde los coeficientes $g_0^{(i)}, g_1^{(i)}, g_2^{(i)}, \ldots, g_M^{(i)}$ son iguales a 0 o 1. De manera correspondiente, el *polinomio generador* de la i-ésima trayectoria está definido por

$$g^{(i)}(D) = g_0^{(i)} + g_1^{(i)}D + g_2^{(i)}D^2 + \cdots + g_M^{(i)}D^M \tag{10.55}$$

donde D denota la variable de retardo unitario. El codificador convolucional completo se describe mediante un conjunto de polinomios generadores $\{g^{(1)}(D), g^{(2)}(D), \ldots, g^{(n)}(D)\}$. Tradicionalmente se emplean diferentes variables para la descripción de los códigos convolucionales y cíclicos, utilizándose comúnmente D en los códigos convolucionales y X en los cíclicos.

▶ EJEMPLO 10.5

Considere el codificador convolucional de la figura 10.13*a*, el cual tiene dos trayectorias numeradas 1 y 2 para una referencia conveniente. La respuesta al impulso de la trayectoria 1 (es decir, la trayectoria superior) es (1, 1, 1). Por tanto, el polinomio generador correspondiente está dado por

$$g^{(1)}(D) = 1 + D + D^2$$

La respuesta al impulso de la trayectoria 2 (es decir, la trayectoria inferior) es (1, 0, 1). De ese modo, el polinomio generador correspondiente se define mediante

$$g^{(2)}(D) = 1 + D^2$$

Para la secuencia de mensaje (10011), por ejemplo, tenemos la representación de polinomio

$$m(D) = 1 + D^3 + D^4$$

Como con la transformada de Fourier, la convolución en el dominio del tiempo se transforma en multiplicación en el dominio D. De esa manera, el polinomio de salida de la trayectoria 1 está dado por

$$\begin{aligned} c^{(1)}(D) &= g^{(1)}(D)m(D) \\ &= (1 + D + D^2)(1 + D^3 + D^4) \\ &= 1 + D + D^2 + D^3 + D^6 \end{aligned}$$

A partir de esto deducimos de inmediato que la secuencia de salida de la trayectoria 1 es (1111001). De modo similar, el polinomio de salida de la trayectoria 2 está dado por

$$\begin{aligned} c^{(2)}(D) &= g^{(2)}(D)m(D) \\ &= (1 + D^2)(1 + D^3 + D^4) \\ &= 1 + D^2 + D^3 + D^4 + D^5 + D^6 \end{aligned}$$

La secuencia de salida de la trayectoria 2 es entonces (1011111). Por último, al multiplexar las dos secuencias de salida de las trayectorias 1 y 2, obtenemos la secuencia codificada

$$\mathbf{c} = (11, 10, 11, 11, 01, 01, 11)$$

Advierta que la secuencia de mensaje de longitud $L = 5$ bits produce una secuencia codificada de longitud $n(L + K - 1) = 14$ bits. Observe también que para que el registro de corrimiento se restablezca a su estado inicial cero, se agrega una secuencia de terminación de $K - 1 = 2$ ceros al último bit de entrada de la secuencia de mensaje. La secuencia de terminación de $K - 1$ ceros recibe el nombre de *cola del mensaje*. ◀

ÁRBOL DE CÓDIGO, ENRAMADO Y DIAGRAMA DE ESTADO

Tradicionalmente, las propiedades estructurales de un codificador convolucional se representan en forma gráfica utilizando uno de tres diagramas equivalentes: árbol de código, enramado y diagrama de estado. Usaremos el codificador convolucional de la figura 10.13*a* como un ejemplo ejecutable para ilustrar la información relevante que cada uno de estos tres diagramas puede ofrecer.

Iniciamos la explicación con el *árbol de código* de la figura 10.14. Cada rama del árbol representa un símbolo de entrada, con el par correspondiente de símbolos binarios de salida indicado sobre la rama. La convención que se utiliza para distinguir los símbolos binarios de entrada 0 y 1 es la siguiente. Una entrada 0 especifica la rama superior de una bifurcación, en tanto que la entrada 1 especifica la rama inferior. Una *trayectoria específica* en el árbol se sigue de izquierda a derecha de acuerdo con la secuencia de entrada (mensaje). Los símbolos codificados correspondientes sobre las ramas de esa trayectoria constituyen la secuencia de entrada (mensaje). Considere, por ejemplo, la secuencia de mensaje (10011) aplicada a la entrada del codificador de la figura 10.13*a*. Siguiendo el procedimiento que acaba de describirse, encontramos que la secuencia codificada correspondiente es (11, 10, 11, 11, 01), la cual coincide con los primeros 5 pares de bits en la secuencia codificada $\{c_i\}$ que se obtuvo en el ejemplo 10.5.

A partir del diagrama de la figura 10.14, observamos que el árbol se vuelve *repetitivo* después de las primeras tres ramas. En realidad, más allá de la tercera rama, los dos nodos marcados con la letra *a* son idénticos, y así son la totalidad de los otros pares de nodos que se marcan de manera idéntica. Es posible establecer esta propiedad repetitiva del árbol examinando el codificador asociado de la figura 10.13*a*. El codificador tiene $M = K - 1 = 2$ bits de mensaje de memoria. Por tanto, cuando el tercer bit de mensaje entra al codificador, el primer bit de mensaje se corre fuera del registro. En consecuencia, después de la tercera rama, la secuencia de mensaje ($100\, m_3 m_4 \ldots$) y ($000\, m_3 m_4 \ldots$) genera los mismos símbolos de código, y puede unirse el par de nodos marcados con la letra *a*. El mismo razonamiento se aplica en los otros nodos. Por tanto, es posible reducir el árbol de código de la figura 10.14 en la nueva forma que se muestra en la figura 10.15, la cual recibe el nombre de un *enramado*.[8] Se llama de esa manera puesto que un enramado es una estructura similar a la de un árbol con ramas que resurgen. La convención utilizada en la figura 10.15 para distinguir entre los símbolos de entrada 0 y 1 es la siguiente. Una rama de código producida por una entrada 0 se dibuja como una línea continua, en tanto que una producida por una entrada 1 se traza como una línea punteada. Como antes, cada secuencia de entrada (mensaje) corresponde a una trayectoria específica a través del enramado. Por ejemplo, vemos de inmediato de la figura 10.15 que la secuencia de mensaje (10011) produce la secuencia de salida codificada (11, 10, 11, 11, 01), lo cual concuerda con nuestro resultado anterior.

Un enramado resulta más instructivo que un árbol en cuanto a que resalta explícitamente el hecho de que el codificador convolucional asociado es una *máquina de estado finito*. Definimos el *estado* de un codificador convolucional de tasa $1/n$ como los $(K - 1)$ bits de mensaje almacenados en el registro de corrimiento del codificador. En el tiempo j, la porción de la secuencia de mensaje que contiene los K bits más recientes se escribe como $(m_{j-K+1}, \ldots, m_{j-1}, m_j)$, donde m_j es el bit *presente*. El estado de $(K - 1)$ bits del codificador en el tiempo j se escribe entonces sim-

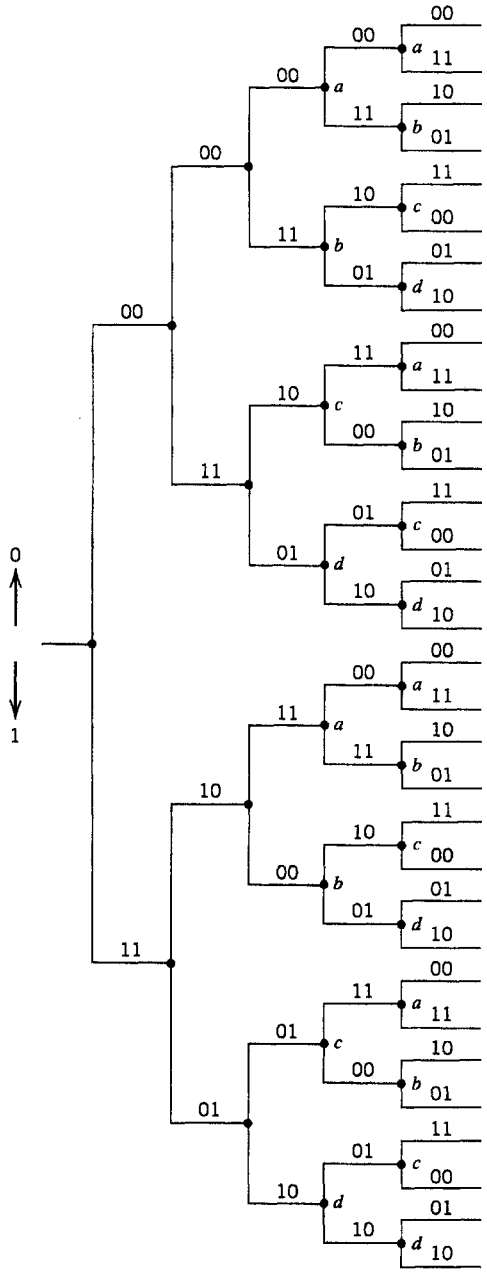


FIGURA 10.14 Árbol de código para el codificador convolucional de la figura 10.13*a*.

plemente como $(m_{j-1}, \ldots, m_{j-K+2}, m_{j-K+1})$. En el caso de un codificador convolucional simple de la figura 10.13*a* tenemos $(K - 1) = 2$. Por tanto, el estado de este codificador puede asumir uno cualquiera de cuatro valores posibles, como se describe en la tabla 10.7. El enramado contiene $(L + K)$ *niveles*, donde L es la longitud de la secuencia de mensaje entrante y K es la longitud de restricción del código. Los niveles del enramado se indican como $j = 0, 1, \ldots, L + K - 1$ en la figura 10.15 para $K = 3$. El nivel j se conoce también como *profundidad* j; ambos términos se emplean de manera indistinta. Los primeros $(K - 1)$ niveles corresponden a la desviación del codificador a partir del estado inicial *a*, y los últimos $(K - 1)$ niveles corresponden al regreso del codificador al estado *a*.

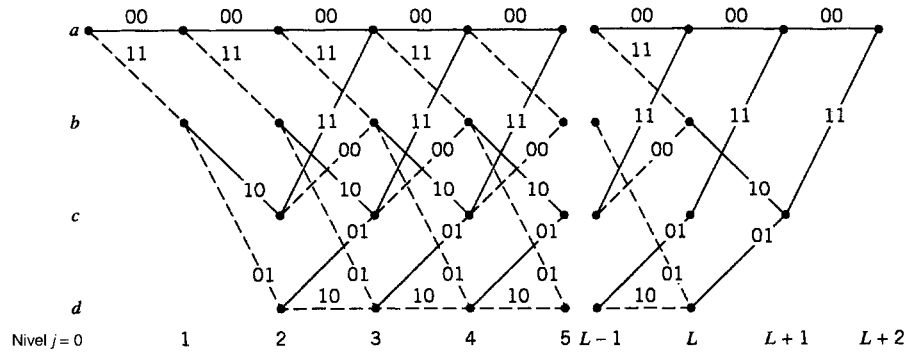


FIGURA 10.15 Árbol para el codificador convolucional de la figura 10.13*a*.

Claramente, no es posible alcanzar todos los estados en estas dos porciones del enramado. Sin embargo, en la parte central del mismo, en el cual el nivel j se encuentra en el intervalo $K - 1 \leq j \leq L$, resultan alcanzables todos los estados del codificador. Advierta también que la parte central del enramado exhibe una estructura periódica fija.

Considere a continuación una parte del enramado correspondiente a los tiempos j y $j + 1$. Suponemos que $j \geq 2$ para el ejemplo que se está considerando, por lo que es posible que el estado presente del codificador sea *a*, *b*, *c* o *d*. Por conveniencia en la presentación, hemos reproducido esta parte del enramado en la figura 10.16*a*. Los nodos de la izquierda representan los cuatro posibles estados presentes del codificador, en tanto que los nodos de la derecha representan los estados siguientes. Claramente, podemos unir los nodos izquierdo y derecho. Al hacerlo, obtenemos el *diagrama de estado* del codificador, ilustrado en la figura 10.16*b*. Los nodos de la figura representan los cuatros estados posibles del codificador, teniendo cada nodo dos ramas entrantes y dos ramas salientes. Una transición de un estado a otro en respuesta a la entrada 0 se representa mediante una rama continua, en tanto que una transición en respuesta a una entrada 1 se ilustra mediante una rama punteada. La marca binaria sobre cada rama representa la salida del codificador cuando ésta se mueve de un estado a otro. Suponga, por ejemplo, que el estado presente del codificador es (01), lo cual se representa por medio del nodo *c*. La aplicación de la entrada 1 al codificador de la figura 10.13*a* produce el estado (10) y la salida codificada (00). Por tanto, con la ayuda de este diagrama de estado, podemos determinar de inmediato la salida del codificador de la figura 10.13*a* para cualquier secuencia de mensaje entrante. Simplemente empezamos en el estado *a*, el estado inicial de puros ceros, y proseguimos a través del diagrama de estado de acuerdo con la secuencia de mensaje. Seguimos una rama continua si la entrada es cero y una rama punteada si ésta es uno. A medida que se recorre cada rama, se genera la marca binaria correspondiente sobre ella. Considere, por ejemplo, la secuencia de mensaje (10011). Para esta entrada seguimos la trayectoria *abcabd* y, por tanto, se genera la secuencia (11, 10, 11, 11, 01), la cual concuerda exactamente con nuestro resultado

TABLA 10.7 ***Tabla de estado para el codificador convolucional de la figura 10.13a***

Estado	*Descripción binaria*
a	00
b	10
c	01
d	11

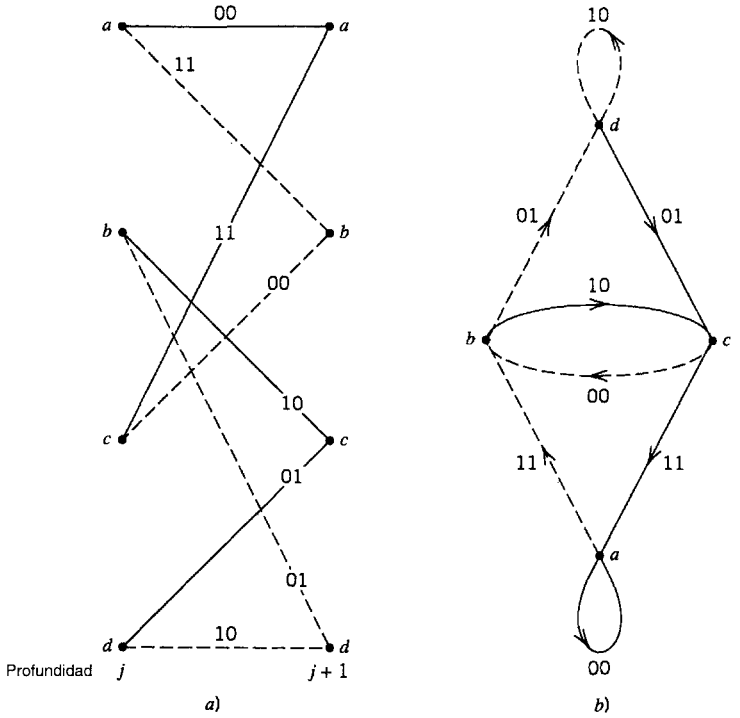


FIGURA 10.16 *a)* Una porción de la parte central del enramado para el codificador de la figura 10.13*a*. *b)* Diagrama de estado del codificador convolucional de la figura 10.13*a*.

anterior. De esa manera, la relación de entrada-salida de un codificador convolucional se describe por completo por medio de su diagrama de estado.

10.6 *Decodificación de máxima verosimilitud de códigos convolucionales*

Ahora que comprendemos la operación de un codificador convolucional, el siguiente aspecto por considerar es la decodificación del mismo. En esta sección describimos primero la teoría fundamental de la decodificación de máxima verosimilitud y presentamos después un algoritmo eficiente para su ejecución práctica.

Sea $\mathbf{m}$ un *vector de mensaje* y $\mathbf{c}$ el correspondiente *vector de código* aplicado por el codificador a la entrada de un canal sin memoria discreto. Sea $\mathbf{r}$ el *vector recibido*, el cual puede diferir del vector de código transmitido debido al ruido de canal. Dado el vector recibido $\mathbf{r}$, se requiere que el decodificador realice una *estimación* $\hat{\mathbf{m}}$ del vector de mensaje. Puesto que existe una correspondencia uno a uno entre el vector de mensaje $\mathbf{m}$ y el vector de código $\mathbf{c}$, el decodificador puede producir de manera equivalente una estimación $\hat{\mathbf{c}}$ del vector de código. En ese caso es posible dejar que $\hat{\mathbf{m}} = \mathbf{m}$ si y sólo si $\hat{\mathbf{c}} = \mathbf{c}$. En otro caso, se comete un *error de decodificación* en el receptor. Se afirma que será óptima la *regla de decodificación* para elegir la estimación $\hat{\mathbf{c}}$, dado el vector recibido $\mathbf{r}$, cuando se minimiza la *probabilidad de error de decodificación*. Con base en el material que se presenta en el capítulo 6, es posible establecer que para mensajes igualmente probables, la probabilidad de error de decodificación se minimiza si la estimación de $\hat{\mathbf{c}}$ se elige para que se maximice la *función de verosimilitud logarítmica*. Sea $p(\mathbf{r}|\mathbf{c})$ la probabilidad condicional de recibir $\mathbf{r}$,

dado que se envió **c**. La función de verosimilitud logarítmica es igual a $\log p(\mathbf{r}|\mathbf{c})$. El *decodificador de máxima verosimilitud* o regla de decisión se describe de la manera siguiente:

$$\begin{gathered}\text{Elíjase la estimación } \hat{\mathbf{c}} \text{ para la cual la función} \\ \text{logarítmica } \log p(\mathbf{r}|\mathbf{c}) \text{ es máxima.}\end{gathered} \tag{10.56}$$

Considere ahora el caso especial de un canal simétrico binario. En este caso, tanto el vector de código transmitido **c** como el vector recibido **r** representan secuencias binarias de longitud N, por ejemplo. Naturalmente, estas dos secuencias quizá difieran entre sí en algunos lugares debido a errores que resultan del ruido del canal. Sean c_i y r_i los elementos i-ésimo de **c** y **r**, respectivamente. Tenemos entonces

$$p(\mathbf{r}|\mathbf{c}) = \prod_{i=1}^{N} p(r_i|c_i) \tag{10.57}$$

De modo correspondiente, la verosimilitud logarítmica es

$$\log p(\mathbf{r}|\mathbf{c}) = \sum_{i=1}^{N} \log p(r_i|c_i) \tag{10.58}$$

Considere que la probabilidad de transición $p(r_i|c_i)$ se define como

$$p(r_i|c_i) = \begin{cases} p, & \text{si } r_i \neq c_i \\ 1-p, & \text{si } r_i = c_i \end{cases} \tag{10.59}$$

Suponga también que el vector **r** difiere del vector de código transmitido **c** en exactamente d posiciones. El número d es la *distancia de Hamming* entre los vectores **r** y **c**. Entonces, es posible reescribir la función de verosimilitud logarítmica en la ecuación (10.58) como

$$\begin{aligned}\log p(\mathbf{r}|\mathbf{c}) &= d \log p + (N-d)\log(1-p) \\ &= d \log\left(\frac{p}{1-p}\right) + N \log(1-p)\end{aligned} \tag{10.60}$$

En general, la probabilidad de que ocurra un error es suficientemente baja para que supongamos que $p < 1/2$. Advertimos también que $N \log(1-p)$ es una constante para toda **c**. Por tanto, podemos reenunciar la regla de decodificación de máxima verosimilitud para el canal simétrico binario de la manera que sigue:

$$\begin{gathered}\text{Elíjase la estimación } \hat{\mathbf{c}} \text{ que minimice la distancia de Hamming} \\ \text{entre el vector recibido } \mathbf{r} \text{ y el vector transmisor } \mathbf{c}.\end{gathered} \tag{10.61}$$

Es decir, para un canal simétrico binario, el decodificador de máxima verosimilitud se reduce a un *decodificador de distancia mínima*. En un decodificador de estas características, el vector recibido **r** se compara con cada vector de código transmitido posible **c** y el particular que se encuentra más próximo a **r** se elige como el vector de código transmitido correcto. El término "más próximo" se utiliza en el sentido del número mínimo de símbolos binarios que difieren (es decir, la distancia de Hamming) entre los vectores de código que se investigan.

El algoritmo de Viterbi[9]

La equivalencia entre la decodificación de máxima verosimilitud y la decodificación de distancia mínima para un canal simétrico binario implica que es posible decodificar un código convolucional eligiendo una trayectoria en el árbol de código cuya secuencia codificada difiere de la secuencia recibida en unos cuantos números de lugares. Puesto que un árbol de código es equivalente a un enramado, es posible igualmente limitar nuestra elección a las trayectorias posibles en la representación de enramado del código. La razón por la que se prefiere el enramado

sobre el árbol es que el número de nodos en cualquier nivel del enramado no continúa creciendo cuando aumenta el número de bits de mensaje entrantes; más bien, se mantiene constante en 2^{K-1}, donde K es la longitud de restricción del código.

Considere, por ejemplo, el diagrama de enramado de la figura 10.15 para un código convolucional con tasa $r = 1/2$ y longitud de restricción $K = 3$. Observamos que en el nivel $j = 3$ existen dos trayectorias que entran a cualquiera de los cuatro nodos en el enramado. Además, estas dos trayectorias serán idénticas hacia adelante desde ese punto. Claramente, un decodificador de distancia mínima quizá tome una decisión en ese punto con respecto a cuál de esas trayectorias retener, sin ninguna pérdida en el desempeño. Es posible tomar una decisión similar en el nivel $j = 4$, y así sucesivamente. Esta secuencia de decisiones es exactamente lo que hace el *algoritmo de Viterbi* cuando avanza a través del enramado. El algoritmo opera al calcular una *métrica* o discrepancia para cada trayectoria posible en el enramado. La métrica para una trayectoria particular se define como la distancia de Hamming entre la secuencia codificada representada por esa trayectoria y la secuencia recibida. Por tanto, en cada nodo (estado) en el enramado de la figura 10.15, el algoritmo compara las dos trayectorias que entran al nodo. Se retiene la trayectoria con métrica inferior y se descarta la otra. Este cálculo se repite para cada nivel j del enramado en el intervalo $M \leq j \leq L$, donde $M = K - 1$ es la memoria del codificador y L es la longitud de la secuencia de mensaje entrante. Las trayectorias que retiene el algoritmo reciben el nombre de *trayectorias sobrevivientes* o *activas*. Para un código convolucional de longitud de restricción $K = 3$, por ejemplo, nunca se almacenarán más de $2^{K-1} = 4$ trayectorias sobrevivientes y sus métricas. Se garantiza siempre que esta lista de 2^{K-1} trayectorias contendrá la elección de máxima verosimilitud.

Una dificultad que puede surgir en la aplicación del algoritmo de Viterbi es la posibilidad de que cuando se comparan las trayectorias que entran a un estado, se encuentra que sus métricas serán idénticas. En una situación tal, nos decidimos por un volado (es decir, hacemos simplemente una suposición).

En resumen, el algoritmo de Viterbi es un decodificador de máxima verosimilitud, que es óptimo para una canal AWGN. Procede paso por paso del modo siguiente:

Asignación de valores iniciales

Se marca como cero el estado más a la izquierda del enramado (es decir, el estado de puros ceros en el nivel 0), pues no hay discrepancia de este punto en el cálculo.

Cálculo del paso $j + 1$

Sea $j = 0, 1, 2, \ldots$, y suponga que en el paso previo j se realizaron dos cosas:

- Se identificaron todas las trayectorias sobrevivientes.
- Se almacenó la trayectoria sobreviviente y su métrica para cada estado del enramado.

En ese caso, en el nivel (tiempo de reloj) $j + 1$, se calcula la métrica para todas las trayectorias que entran en cada estado del enramado agregando la métrica de las ramas entrantes a la métrica de la trayectoria sobreviviente conectora desde el nivel j. Por tanto, en cada estado se identifica la trayectoria con la métrica más baja como la sobreviviente del paso $j + 1$, actualizando de esa manera el cálculo.

Paso final

Continúa el cálculo hasta que el algoritmo completa su búsqueda hacia adelante, a través del enramado y, por tanto, llega al nodo de terminación (es decir, el estado de puros ceros), momento en el que toma una decisión acerca de la trayectoria de máxima verosimilitud. Entonces, al igual que un decodificador de bloque, la secuencia de símbolos asociada con esa trayectoria se libera hacia el destino como la versión decodificada de la secuencia recibida. En este sentido, resulta entonces más correcto referirse al algoritmo de Viterbi como un *estimador de secuencia de máxima verosimilitud*.

Sin embargo, cuando la secuencia recibida es muy larga (casi infinita), el requerimiento de almacenamiento del algoritmo de Viterbi se vuelve demasiado alto y deben establecerse algunos compromisos. El método que suele seguirse consiste en "truncar" la memoria de la trayectoria del decodificador, como se describe a continuación. Una *ventana de decodificación* de longitud l se especifica y el algoritmo opera sobre una trama correspondiente de la secuencia recibida, interrumpiéndose siempre después de l pasos. Se realiza entonces una decisión con respecto a la "mejor" trayectoria y se libera al usuario el símbolo asociado con la primera rama sobre esa trayectoria. Se elimina el símbolo asociado con la última rama de la trayectoria. A continuación, la ventana de decodificación se mueve hacia adelante un intervalo de tiempo, y se toma una decisión sobre la siguiente trama de código, etc. Las decisiones de decodificación que se llevan a cabo de esta manera ya no son en verdad de máxima verosimilitud, aunque pueden hacerse casi tan buenas siempre que la ventana de decodificación sea suficientemente larga. La experiencia y el análisis han demostrado que se obtienen resultados satisfactorios si la longitud de la ventana de decodificación l es del orden de 5 veces la longitud de restricción K del código convolucional o mayor.

Ejemplo 10.6 Decodificación correcta de la secuencia de puros ceros recibida

Suponga que el codificador de la figura 10.13*a* genera una secuencia de puros ceros que se envía por un canal simétrico binario, y que la secuencia recibida es (0100010000...). Hay dos errores en la secuencia recibida debido al ruido en el canal: uno en el segundo bit y el otro en el sexto bit. Deseamos demostrar que este patrón de error doble es corregible mediante la aplicación del algoritmo de decodificación de Viterbi.

En la figura 10.17 presentamos los resultados de la aplicación del algoritmo para el nivel $j = 1, 2, 3, 4, 5$. Vemos que para $j = 2$ hay (para la primera vez) cuatro trayectorias, una para cada uno de los cuatro estados del codificador. La figura incluye también la métrica de cada trayectoria para cada nivel en el cálculo.

En el lado izquierdo de la figura 10.17, para $j = 3$ mostramos las trayectorias que entran a cada uno de los estados, junto con sus métricas individuales. En el lado derecho de la figura ilustramos las cuatro sobrevivientes que resultan de la aplicación del algoritmo en el nivel $j = 3, 4, 5$.

Al examinar las cuatro sobrevivientes de la figura 10.17 para $j = 5$, vemos que la trayectoria de puros ceros tiene la métrica más pequeña y seguirá siendo la trayectoria de métrica menor desde este punto hacia adelante. Esto demuestra claramente que la secuencia de puros ceros es la elección de máxima verosimilitud del algoritmo de decodificación de Viterbi, la cual concuerda exactamente con la secuencia transmitida. ◀

Ejemplo 10.7 Decodificación incorrecta de la secuencia de puros ceros recibida

Suponga a continuación que la secuencia recibida es (1100010000...), la cual contiene tres errores al compararse con la secuencia de puros ceros transmitida.

En la figura 10.18 mostramos los resultados de la aplicación del algoritmo de decodificación de Viterbi para $j = 1, 2, 3, 4$. Vemos que en este ejemplo la trayectoria correcta se ha eliminado en el nivel $j = 3$. Es claro que un patrón de tres errores resulta incorregible mediante el algoritmo de Viterbi cuando se aplica a un código convolucional de tasa 1/2 y longitud de restricción $K = 3$. La excepción a esta regla es un patrón de error triple distribuido por un periodo más largo que una longitud de restricción, en cuyo caso es muy probable que sea corregible. ◀

Distancia libre de un código convolucional

El desempeño de un código convolucional depende no sólo del algoritmo de decodificación utilizado, sino también de las propiedades de distancia del código. En este contexto, la medida simple más importante de la capacidad de un código convolucional para combatir el ruido del canal es la distancia libre, denotada por d_{libre}. La *distancia libre* de un código convolucional se define como *la distancia de Hamming mínima entre cualesquiera dos palabras de código en el código*. Un código convolucional con distancia libre d_{libre} puede corregir t errores si y sólo si d_{libre} es mayor que $2t$.

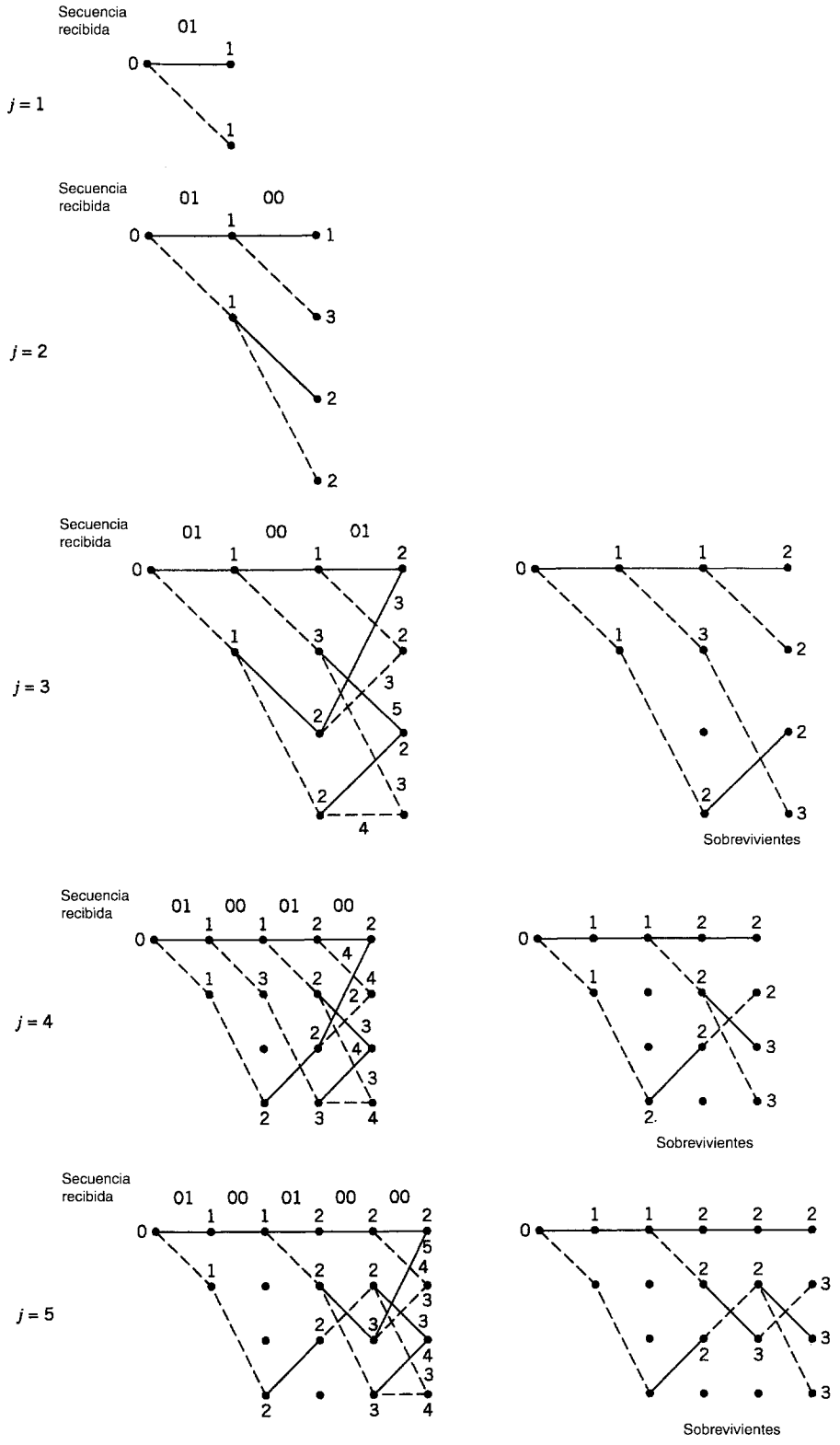


FIGURA 10.17 Ilustración de los pasos en el algoritmo de Viterbi para el ejemplo 10.6.

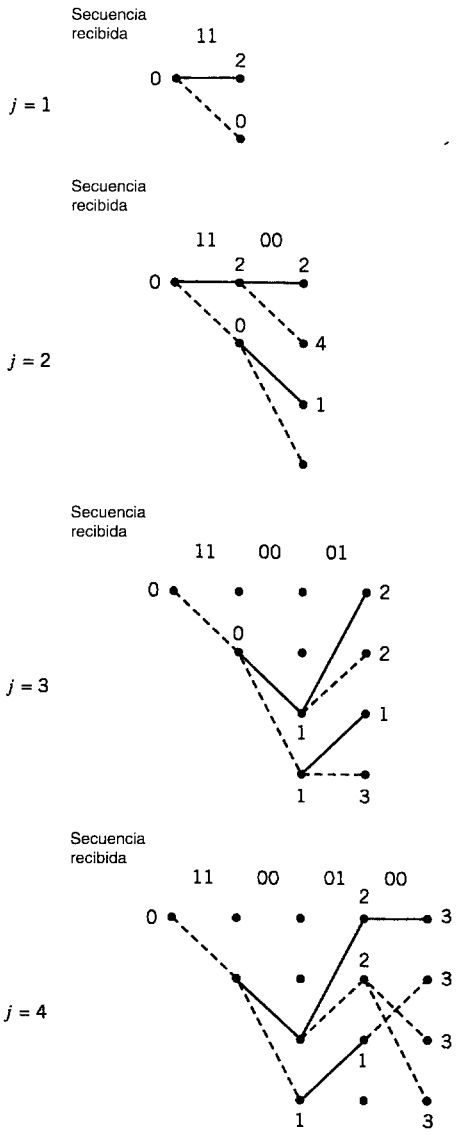


FIGURA 10.18 Ilustración de la ruptura del algoritmo de Viterbi en el ejemplo 10.7.

La distancia libre puede obtenerse en forma bastante simple a partir del diagrama de estado del codificador convolucional. Considere, por ejemplo, la figura 10.16*b*, la cual muestra el diagrama de estado del codificador de la figura 10.13*a*. Cualquier secuencia de código distinta de cero corresponde a una trayectoria completa que empieza y termina en el estado 00 (es decir, el nodo *a*). Encontramos entonces de utilidad dividir este nodo de la manera que se ilustra en el diagrama de estado modificado de la figura 10.19, el cual puede considerarse como una *gráfica del flujo de señal* con una sola entrada y una sola salida. Una gráfica del flujo de señal consiste en *nodos* y *ramas* dirigidas; opera mediante las siguientes reglas:

1. Una rama multiplica la señal en su nodo de entrada por la *transmitancia* que caracteriza a esa rama.
2. Un nodo con ramas entrantes *suma* las señales producidas por todas esas ramas.
3. La señal en un nodo se aplica igualmente a todas las ramas que salen de ese nodo.

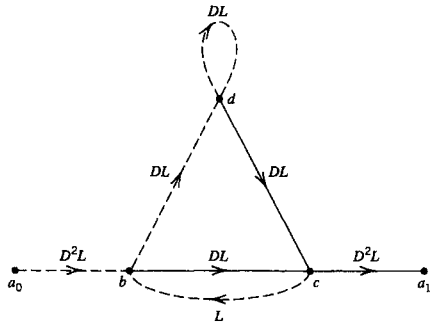


FIGURA 10.19 Diagrama de estado modificado del codificador convolucional.

4. La *función de transferencia* de la gráfica es la proporción entre la señal de salida y la señal de entrada.

Volviendo a la gráfica del flujo de señales de la figura 10.19, observamos que el exponente de D sobre una rama en esta gráfica describe el peso de Hamming de la salida del codificador correspondiente a esa rama. El exponente de L es siempre igual a uno, pues la longitud de cada rama es uno. Sea $T(D, L)$ la función de transferencia de la gráfica del flujo de señal, con D y L desempeñando el papel de variables simuladas. Para el ejemplo de la figura 10.19, es posible utilizar de inmediato las reglas 1, 2 y 3 para obtener las siguientes relaciones de entrada-salida:

$$\left.\begin{aligned} b &= D^2La_0 + Lc \\ c &= DLb + DLd \\ d &= DLb + DLd \\ a_1 &= D^2Lc \end{aligned}\right\} \tag{10.62}$$

donde a_0, b, c, d y a_1 denotan las señales de nodo de la gráfica. Resolviendo este conjunto de ecuaciones (10.62) para la proporción a_1/a_0, encontramos que la función de transferencia de la gráfica en la figura 10.19 está dada por

$$T(D, L) = \frac{D^5L^3}{1 - DL(1 + L)} \tag{10.63}$$

Utilizando la expansión del binomio, es posible escribir de manera equivalente

$$T(D, L) = D^5L^3 \sum_{i=0}^{\infty} (DL(1 + L))^i \tag{10.64}$$

Dejando $L = 1$ en la ecuación (10.64), obtenemos entonces la *función de transferencia de distancia* expresada en la forma de una serie de potencias:

$$T(D, 1) = D^5 + 2D^6 + 4D^7 + \ldots \tag{10.65}$$

Puesto que la distancia libre es la distancia de Hamming mínima entre cualesquiera dos palabras de código en el código y la función de transferencia de distancia $T(D, 1)$ enumera el número de palabras de código que están a una distancia de separación determinada, se concluye que el exponente del primer término en la expansión de $T(D, 1)$ define la distancia libre. De ese modo, con base en la ecuación (10.65), el código convolucional de la figura 10.13*a* tiene una distancia libre $d_{\text{libre}} = 5$.

Este resultado indica que hasta dos errores en la secuencia recibida son corregibles, ya que dos o un número menor de errores de transmisión provocarán que la secuencia recibida se encuentre a lo más a una distancia de Hamming de 2 de la secuencia transmitida, aunque a

TABLA 10.8 ***Máximas distancias libres alcanzables con códigos convolucionales sistemáticos y no sistemáticos de tasa 1/2***

Longitud restrictiva K	*Sistemático*	*No sistemático*
2	3	3
3	4	5
4	4	6
5	5	7
6	6	8
7	6	10
8	7	10

menos de una distancia de Hamming de 3 desde cualquier otra secuencia de código. En otras palabras, a pesar de la presencia de cualquier par de errores de transmisión, la secuencia recibida permanece más cercana a la secuencia transmitida que cualquier otra posible secuencia de código. Sin embargo, este enunciado deja de ser válido si hay tres o más errores de *transmisión espaciados a muy corta distancia* en la secuencia recibida. Estas observaciones confirman los resultados que se indicaron antes en los ejemplos 10.6 y 10.7.

Al utilizar la función de transferencia de distancia $T(D, 1)$ para calcular la distancia libre de un código convolucional, se supone que la serie de potencias en la variable de retardo unitario D que representa a $T(D, 1)$ es *convergente* (es decir, su suma tiene un valor "finito"). Esta suposición se necesita para justificar la expansión que se da en la ecuación (10.65) para el código convolucional de la figura 10.13*a*. Sin embargo, no hay garantía de que $T(D, 1)$ sea siempre convergente. Cuando $T(D, 1)$ es no convergente, se ocasiona un número infinito de errores de decodificación mediante un número finito de errores de transmisión; el código convolucional está sujeto entonces a una propagación de errores catastrófica, y el código recibe el nombre de *código catastrófico*.[10] En este contexto vale la pena señalar que un código convolucional *sistemático* no puede ser catastrófico. Desafortunadamente, para una longitud de restricción preestablecida K, las distancias libres que pueden conseguirse con códigos convolucionales sistemáticos utilizando esquemas como el que se muestra en la figura 10.13, suelen ser más pequeños que en el caso de códigos convolucionales no sistemáticos, como se indica en la tabla 10.8.

GANANCIA DE CODIFICACIÓN ASINTÓTICA[11]

La función de transferencia del diagrama de estado del codificador, modificada de una manera similar a la que se ilustra en la figura 10.19, puede utilizarse para evaluar *una cota sobre la tasa de error de bits* para un esquema de decodificación dado; sin embargo, los detalles de esta evaluación están más allá del objetivo de nuestro análisis presente. Aquí resumimos simplemente los resultados de dos canales especiales, a saber, el canal simétrico binario y el canal con ruido blanco gaussiano aditivo (AWGN) de entrada binaria, suponiendo el uso de corrimiento de fase por llaveo (PSK) binaria con detección coherente.

1. *Canal simétrico binario.* Éste puede modelarse como un canal con ruido blanco gaussiano aditivo con corrimiento de fase por llaveo (PSK) binaria como la modulación y con demodulación de decisiones rígidas. La probabilidad de transición p del canal simétrico binario es entonces igual a la tasa de error de bits (BER) para el sistema PSK binario no codificado. Del capítulo 6 recordamos que para valores grandes de E_b/N_0, la razón entre la energía de la señal por bit y la densidad espectral de potencia del ruido, el factor exponencial $\exp(-E_b/N_0)$ domina la tasa de error de bits correspondiente a la PSK binaria sin codificación. Por otra parte, la tasa de error de bits para el mismo esquema de modulación con codificación convolucional resulta dominada por el factor exponencial $\exp(-d_{\text{libre}} rE_b/2N_0)$, donde r es la tasa de código y d_{libre} es la distancia libre del código

convolucional. En consecuencia, como una figura de mérito para medir la mejora en el desempeño con respecto a los errores que se cometen por el uso de codificación con decodificación de decisiones rígidas, podemos utilizar los exponentes para definir la *ganancia de codificación asintótica* (en decibeles) como sigue:

$$G_a = 10 \log_{10}\left(\frac{d_{\text{libre}}}{2}\right) \text{dB} \tag{10.66}$$

2. *Canal AWGN de entrada binaria.* Considere a continuación el caso de un canal AWGN sin memoria de entrada binaria y sin cuantización de salida [es decir, la amplitud de salida se encuentra en el intervalo $(-\infty, \infty)$]. Para este canal, la teoría demuestra que para valores grandes de E_b/N_0, la tasa de error de bits correspondientes a la PSK binaria con codificación convolucional está dominada por el factor exponencial $\exp(-d_{\text{libre}} r E_b/N_0)$, donde los parámetros son como se definieron antes. En este caso encontramos que la ganancia de codificación asintótica está definida por

$$G_a = 10 \log_{10}(d_{\text{libre}} r) \text{dB} \tag{10.67}$$

A partir de las ecuaciones (10.66) y (10.67) vemos que la ganancia de codificación asintótica para el canal AWGN de entrada binaria es mayor que la correspondiente al canal simétrico binario en 3 dB. En otras palabras, para E_b/N_0 grande, el transmisor de un canal simétrico binario debe generar 3 dB adicionales de energía (o potencia) de la señal sobre la que corresponde al canal AWGN de entrada binaria si se va a conseguir el mismo desempeño en relación con los errores. Claramente existe una ventaja que se logra al permitir una salida de demodulador no cuantizada en lugar de realizar decisiones rígidas. Sin embargo, esta mejora en el desempeño se alcanza a costa de la mayor complejidad del decodificador debida al requerimiento de aceptar entradas analógicas.

La ganancia de codificación asintótica correspondiente a un canal AWGN de entrada binaria se aproxima hasta dentro de aproximadamente 0.25 dB mediante un canal sin memoria discreto de salida Q-aria y entrada binaria con el número de niveles de representación $Q = 8$. Esto quiere decir que es posible evitar la necesidad de un decodificador analógico utilizando un decodificador de decisiones flexibles que realiza la cuantización de salida finita (por lo común, $Q = 8$), y llevar a cabo un desempeño próximo al óptimo.

10.7 *Modulación por codificación de enramado*[12]

En el método tradicional relativo a la codificación de canal que se describe en las secciones anteriores del capítulo, la codificación se efectúa independientemente de la modulación en el transmisor; lo mismo sucede con respecto a la decodificación y detección en el receptor. Además, se proporciona control de errores al transmitir bits redundantes adicionales en el código, lo cual tiene el efecto de disminuir la tasa de bits de información por ancho de banda del canal. Es decir, la eficiencia del ancho de banda se intercambia por una mayor eficiencia de potencia.

Para alcanzar una utilización más efectiva del ancho de banda y la potencia disponibles, la codificación y la modulación tienen que considerarse como una sola entidad. Es posible abordar esta nueva situación redefiniendo la codificación como *el proceso de imponer ciertos patrones sobre la señal transmitida.* En realidad, esta definición incluye la idea tradicional de codificación de paridad.

Los *códigos de enramado* para los canales limitados en banda resultan del tratamiento de la modulación y la codificación como una entidad *combinada* en vez de dos operaciones separadas. La propia combinación se conoce como *modulación por codificación de enramado* (TCM). Esta forma de transmisión de señales tiene tres rasgos básicos:

1. El número de puntos de señal en la constelación utilizada es mayor que el requerido en el formato de modulación de interés con la misma tasa de datos; los puntos adicionales permiten la redundancia para la codificación directa del control de error sin sacrificar ancho de banda.

2. La codificación convolucional se utiliza para introducir cierta independencia entre los puntos de señal sucesivos, de tal modo que sólo se permiten ciertos *patrones* o *secuencias de puntos de señal*.
3. La decodificación de decisiones flexibles se realiza en el receptor, en el cual la secuencia permisible de señales se modela como una estructura de enramado; de ahí el nombre de "códigos de enramado".

Este último requerimiento es el resultado de utilizar una constelación de señales agrandada. Al incrementar el tamaño de la constelación, aumenta la probabilidad del error de símbolo para una relación de señal a ruido fija. Por tanto, con la demodulación de decisiones rígidas antes de empezar encararíamos una pérdida de desempeño. La realización de la decodificación de decisiones flexibles sobre el enramado de codificación y modulación combinado alivia este problema.

En la presencia del AWGN, la decodificación de máxima verosimilitud de los códigos de enramado consiste en determinar aquella trayectoria particular a través del enramado *con distancia euclidiana cuadrática mínima* para la secuencia recibida. Por lo tanto, en el diseño de los códigos de enramado, el énfasis radica en maximizar la distancia euclidiana entre los vectores de código (o, equivalentemente, las palabras de código) en vez de maximizar la distancia de Hamming de un código de corrección de errores. La razón de este procedimiento es que, salvo para la codificación convencional con PSK binaria y QPSK, la maximización de la distancia de Hamming no es lo mismo que la maximización de la distancia euclidiana al cuadrado. En seguida se adopta la distancia euclidiana como la medida de distancia de interés. Además, si bien es posible un tratamiento más general, el análisis (por elección) se confina al caso de *constelaciones bidimensionales de los puntos de la señal*. La implicación de una elección de este tipo es restringir el desarrollo de códigos de enramado a esquemas de modulación por amplitud y/o fase multinivel, tales como la PSK M-aria y la QAM M-aria.

El método utilizado para diseñar este tipo de códigos de enramado implica dividir una constelación M-aria de interés sucesivamente en 2, 4, 8, ... subconjuntos con tamaño M/2, M/4, M/8, ..., y en tener progresivamente mayores distancias euclidianas mínimas crecientes entre sus puntos de señal respectivos. Un método de diseño como éste mediante la *división de conjunto* representa la "idea clave" en la construcción de técnicas de modulación de codificación eficientes para canales limitados en banda.

En la figura 10.20 ilustramos el procedimiento de división considerando una constelación circular que corresponde a la PSK 8. La figura describe la propia constelación y los subconjuntos

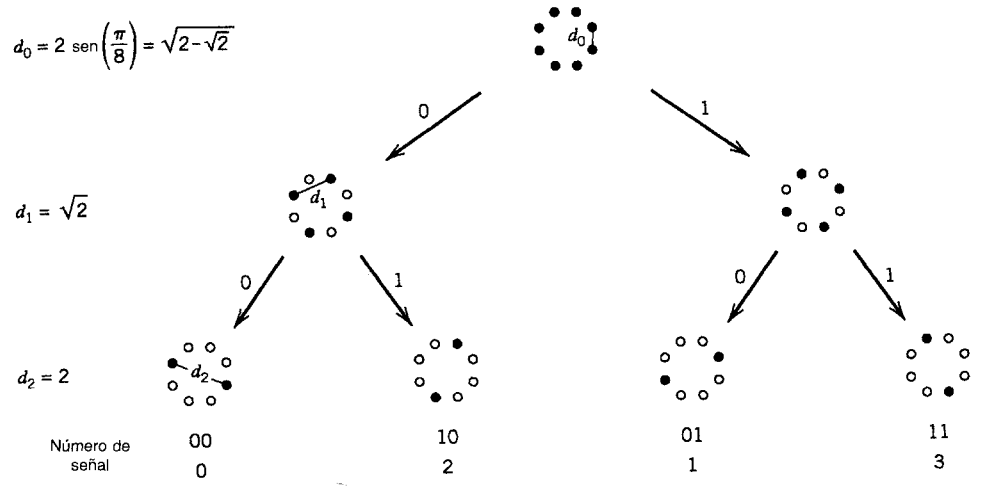


FIGURA 10.20 División de la constelación PSK 8, la cual muestra que $d_0 < d_1 < d_2$.

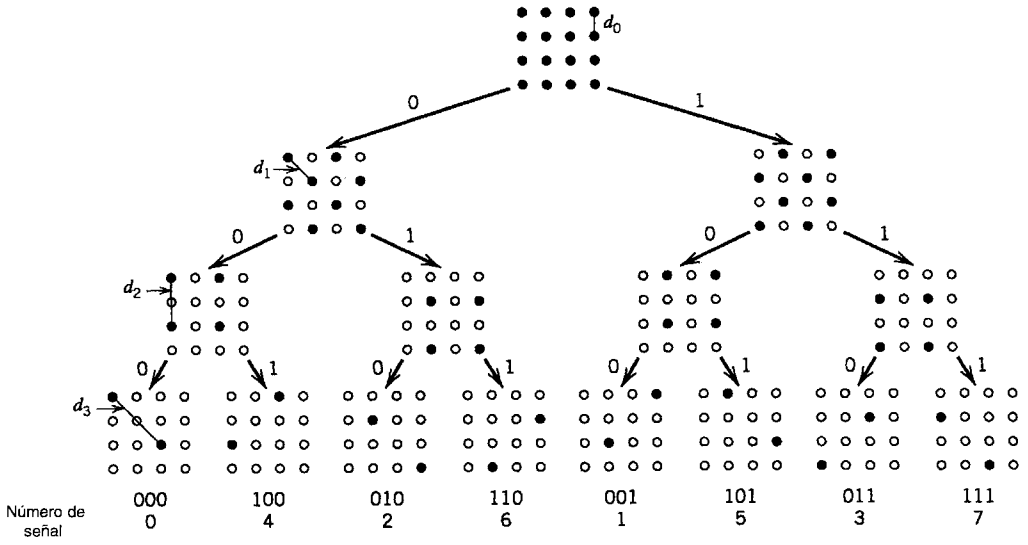


FIGURA 10.21 División de la constelación QAM-16, la cual muestra que $d_0 < d_1 < d_2 < d_3$.

2 y 4 que resultan de dos niveles de división. Estos subconjuntos comparten la propiedad común de que las distancias euclidianas mínimas entre sus puntos individuales siguen un patrón creciente: $d_0 < d_1 < d_2$.

La figura 10.21 ilustra la división de una constelación rectangular correspondiente a la QAM 16. Aquí también vemos que los subconjuntos tienen un incremento dentro de las distancias euclidianas del subconjunto: $d_0 < d_1 < d_2 < d_3$.

Con base en los subconjuntos que resultan de la división sucesiva de una constelación bidimensional, es posible idear esquemas de codificación relativamente simples e incluso altamente efectivos. Específicamente, para enviar n bits/símbolo con *modulación en cuadratura* (es decir, una que tiene componentes en fase y en cuadratura), empezamos con una constelación bidimensional de 2^{n+1} puntos de señal apropiada para el formato de modulación de interés; una malla circular se utiliza para la PSK M-aria, y una rectangular para la QAM M-aria. En cualquier caso, la constelación se divide en subconjuntos de 4 u 8. Uno o dos bits entrantes por símbolo ingresan a un codificador convolucional binario de tasa 1/2 o 2/3, respectivamente; los dos o tres bits codificados por símbolo que resultan determinan la elección del subconjunto particular. Los bits de datos no codificados que quedan determinan qué punto particular del subconjunto elegido se va a transmitir por medio de señales. Esta clase de códigos de enramado se conocen como *códigos Ungerboeck*.

Puesto que el modulador tiene memoria, recurrimos al algoritmo de Viterbi para realizar la estimación de la secuencia de máxima verosimilitud en el receptor. Cada rama en el enramado del código Ungerboeck corresponde a un subconjunto más que a un punto de la señal individual. El primer paso en la detección implica determinar el punto de la señal dentro de cada subconjunto que es más próximo al punto de la señal recibida en el sentido euclidiano. El punto de la señal determinado de esta manera y su métrica (es decir, la distancia euclidiana al cuadrado entre él y el punto recibido) puede utilizarse de aquí en adelante para la rama en cuestión, y el algoritmo de Viterbi puede continuar entonces de la manera usual.

■ CÓDIGOS UNGERBOECK PARA LA PSK 8

El esquema de la figura 10.22*a* describe al código Ungerboeck PSK 8 más simple para la transmisión de 2 bits/símbolo. El esquema utiliza un codificador convolucional de tasa 1/2; el enramado

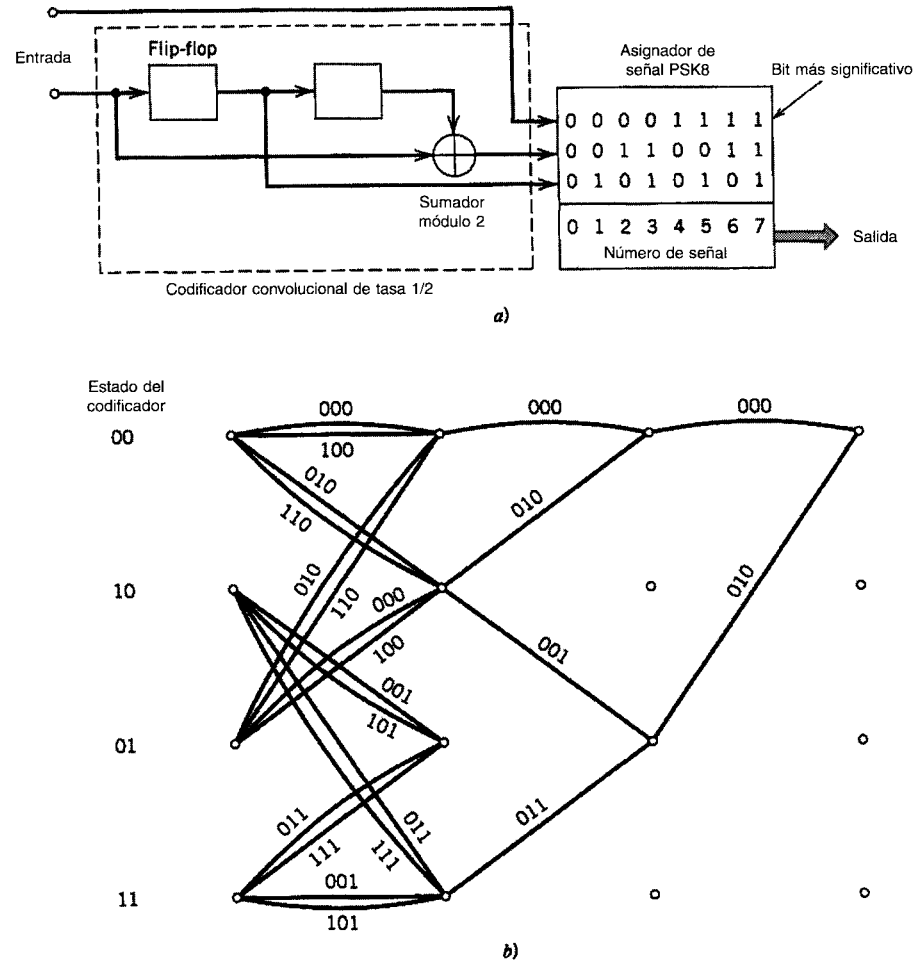
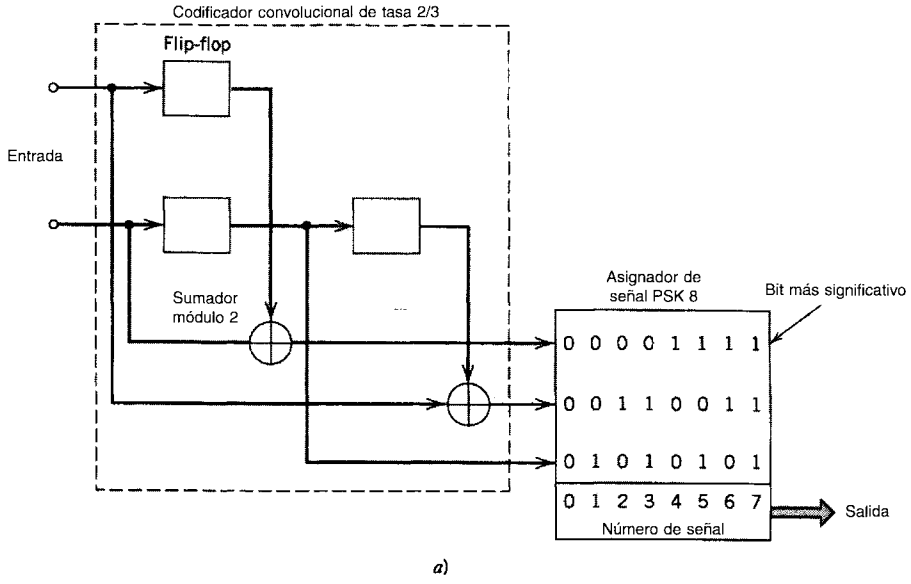


FIGURA 10.22 *a*) Código Ungerboeck de cuatro estados para la PSK 8; el asignador se obtiene de la figura 10.20. *b*) Enramado del código.

correspondiente del código se presenta en la figura 10.22*b*, el cual tiene cuatro estados. Advierta que el bit más significativo de la palabra binaria entrante se deja sin codificar. Por tanto, cada rama del enramado puede corresponder a dos valores de salida diferentes del modulador PSK 8 o, equivalentemente, a uno de los cuatro subconjuntos de 2 puntos que se muestra en la figura 10.20. El enramado de la figura 10.22*b* incluye también la trayectoria de distancia mínima.

El esquema de la figura 10.23*a* describe otro código Ungerboeck PSK 8 para transmitir 2 bits/muestra; éste es el siguiente en nivel de complejidad. El segundo esquema utiliza un codificador convolucional de tasa 2/3. Por tanto, el enramado correspondiente del código tiene ocho estados, como se indica en la figura 10.23*b*. En este caso, ambos bits de la palabra binaria entrante se codifican. Por tanto, cada rama del enramado corresponde a un valor de salida específico del modulador PSK 8. El enramado de la figura 10.23*b* incluye también la trayectoria de distancia mínima.

Las figuras 10.22*b* y 10.23*b* incluyen también los estados del codificador. En la figura 10.22, el estado del codificador se define por medio de los contenidos del registro de corrimiento de dos etapas. Por otra parte, en la figura 10.23 se define por medio del contenido del registro de corrimiento de una etapa (parte superior), seguido por el del registro de corrimiento de dos etapas (parte inferior).



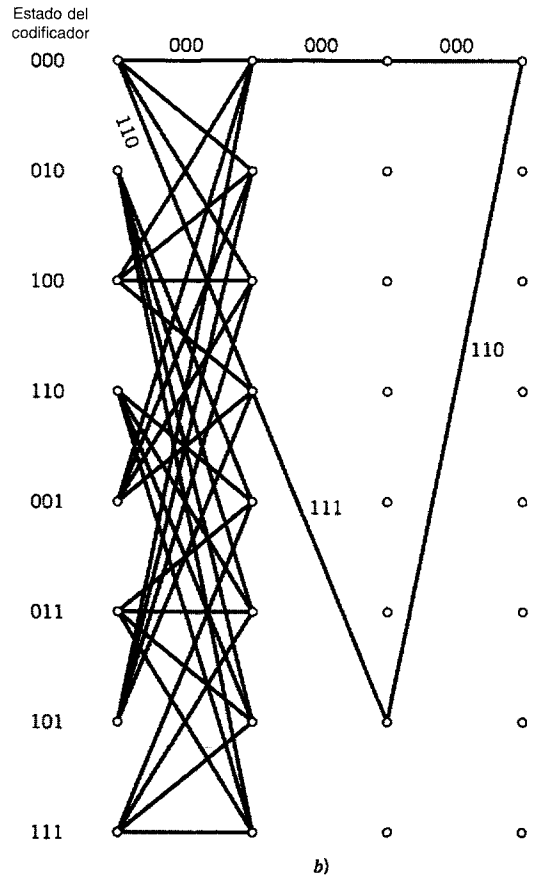


FIGURA 10.23 *a*) Código Ungerboeck de ocho estados para la PSK 8; el asignador se obtiene de la figura 10.20. *b*) Enramado del código con sólo algunas de las ramas mostradas.

Ganancia de codificación asintótica

Siguiendo la explicación en la sección 10.6, definimos la *ganancia de codificación asintótica* de los códigos Ungerboeck como

$$G_a = 10 \log_{10}\left(\frac{d^2_{\text{libre}}}{d^2_{\text{ref}}}\right) \tag{10.68}$$

donde d_{libre} es la *distancia euclidiana libre* del código y d_{ref} es la distancia euclidiana mínima de un esquema de modulación sin codificación que opera con la misma energía de señal por bit. Por ejemplo, al usar el código PSK 8 Ungerboeck de la figura 10.22*a*, la constelación de señales tiene 8 puntos de mensaje y enviamos 2 bits de mensaje por punto. Por tanto, la transmisión sin codificación requiere una constelación de señales con 4 puntos de mensaje. Es posible considerar entonces la PSK 4 sin codificación como la referencia para el código PSK 8 Ungerboeck de la figura 10.22*a*.

El código PSK 8 Ungerboeck de la figura 10.22*a* alcanza una ganancia de codificación asintótica de 3 dB, que se calcula de la manera siguiente:

1. Cada rama del enramado de la figura 10.22*b* corresponde a un subconjunto de dos puntos de señal antípoda. De ese modo, la distancia euclidiana libre d_{libre} del código puede no ser mayor que la distancia euclidiana d_2 entre los puntos de la señal antípoda de un subconjunto de este tipo. Es posible escribir entonces

$$d_{\text{libre}} = d_2 = 2$$

   donde la distancia d_2 se define en la figura 10.24*a*; vea también la figura 10.20.
2. La distancia euclidiana mínima de una QPSK sin codificación, vista como una operación de referencia con la misma energía de señal por bit, es igual a (vea la figura 10.24*b*)

$$d_{\text{ref}} = \sqrt{2}$$

Por tanto, como se estableció antes, el uso de la ecuación (10.68) produce una ganancia de codificación asintótica de $10 \log_{10} 2 = 3$ dB.

La ganancia de codificación asintótica alcanzable con los códigos Ungerboeck aumenta con el número de estados en el codificador convolucional. La tabla 10.9 presenta la ganancia de codificación asintótica (en dB) para los códigos PSK 8 Ungerboeck correspondientes a un nú-

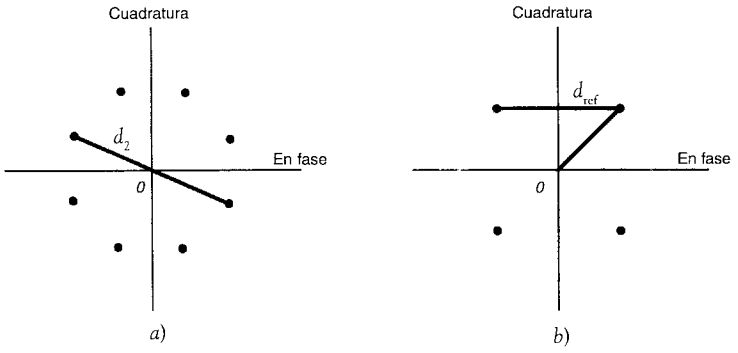


FIGURA 10.24 Diagramas del espacio de la señal para el cálculo de la ganancia de codificación asintótica del código Ungerboeck PSK 8. *a*) Definición de la distancia d_2. *b*) Definición de la distancia de referencia d_{ref}.

TABLA 10.9 *Ganancia de codificación asintótica de los códigos PSK8 Ungerboeck, con respecto a la PSK4 sin codificación*

Número de estados	4	8	16	32	64	128	256	512
Ganancia de codificación (dB)	3	3.6	4.1	4.6	4.8	5	5.4	5.7

mero creciente de estados, expresados con respecto a la PSK4 sin codificar. Advierta que las mejoras del orden de 6 dB requieren códigos con un número muy grande de estados.

10.8 *Códigos turbo*[13]

Tradicionalmente, el diseño de buenos códigos se ha abordado construyendo códigos con una gran estructura algebraica, para los cuales hay esquemas de decodificación factibles. Un procedimiento de este tipo se ejemplifica mediante los códigos de bloque lineales y los códigos convolucionales que se explican en las secciones anteriores. La dificultad con estos códigos tradicionales radica en que, con el fin de alcanzar el límite teórico de la capacidad del canal de Shannon, necesitamos aumentar la longitud de la palabra de un código de bloques lineal o la longitud de restricción de un código convolucional, lo cual, a su vez, provoca que la complejidad de cómputo de un codificador de máxima verosimilitud aumente de manera exponencial. A la larga, llegamos a un punto donde la complejidad del decodificador es tan alta que éste se vuelve físicamente irrealizable.

Se han propuesto varios métodos para la construcción de códigos poderosos con grandes longitudes de bloque "equivalentes" estructurados de tal manera que la decodificación puede dividirse en varias etapas manejables. Con base en los métodos anteriores, el desarrollo de los *códigos turbo* y de los *códigos de verificación de paridad de baja densidad* ha sido por mucho el de mayor éxito. En realidad este desarrollo ha abierto una nueva rama y una manera excitante de construir códigos adecuados y de decodificarlos con complejidad factible. Los códigos turbo se explican en esta sección y los códigos de verificación de paridad de baja densidad se presentan en la sección 10.10.

CODIFICACIÓN TURBO

En su forma más básica, el codificador de un código turbo consiste en dos codificadores sistemáticos *constituyentes* unidos por medio de un intercalamiento, como se ilustra en la figura 10.25.

Un *intercalamiento* es un dispositivo de asignación de entrada-salida que *permuta* el orden de una secuencia de símbolos proveniente de un alfabeto fijo de una manera completamente determinística; es decir, toma los símbolos a la entrada y produce símbolos idénticos a la salida, pero en un orden temporal diferente. El intercalador puede ser de muchos tipos, dos de los cuales son el periódico y el pseudoaleatorio. Los códigos turbo utilizan el intercalador pseudoaleatorio, el

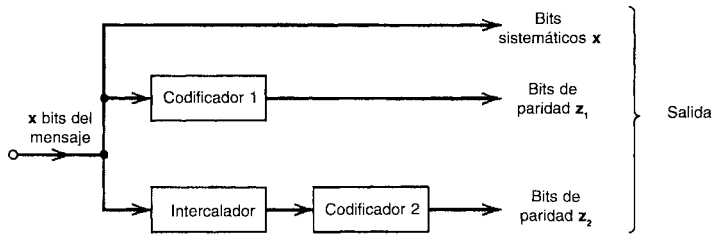


FIGURA 10.25 Diagrama de bloques de un codificador turbo.

cual opera únicamente sobre los bits sistemáticos. Son dos las razones del uso de un intercalador en un código turbo:

- Para ligar errores que se efectúan con facilidad en una mitad del código turbo con errores cuya ocurrencia resulta excepcionalmente improbable en la otra mitad. Ésta es en realidad la razón principal por la que el código turbo se desempeña mejor que un código tradicional.
- Para ofrecer un desempeño adecuado con respecto a la decodificación sin acoplamiento, la cual es un problema que se presenta cuando las estadísticas del canal no se conocen o se han especificado incorrectamente.

Por lo común, aunque no necesariamente, se utiliza el mismo código para ambos codificadores constituyentes en la figura 10.25. Los códigos constituyentes que se recomiendan para los códigos turbo son los *convolucionales sistemáticos recursivos (RSC)* de longitud de restricción corta. La razón por la que estos códigos convolucionales se hacen recursivos (es decir, alimentan una o más de las salidas de derivación en el registro de corrimiento de regreso a la entrada) radica en hacer que el estado interno del registro de corrimiento dependa de las salidas pasadas. Lo anterior afecta el comportamiento de los patrones de error (un error simple en los bits sistemáticos produce un número infinito de errores de paridad), resultando en que se logra un mejor desempeño de la estrategia de codificación total.

EJEMPLO 10.8 Codificador RSC de ocho estados

La figura 10.26 muestra un ejemplo de codificador RSC de ocho estados. La matriz generadora correspondiente a este código convolucional recursivo es

$$g(D) = \left[1, \frac{1 + D + D^2 + D^3}{1 + D + D^3}\right] \tag{10.69}$$

donde D es la variable de retardo. La segunda entrada de la matriz $\mathbf{g}(D)$ es la función de transferencia del registro de corrimiento realimentado, que se define como la transformada de la salida dividida entre la transformada de la entrada. Sea $M(D)$ la transformada de la secuencia de mensaje $\{m_i\}_{i=1}^{k}$ y $B(D)$ la transformada de la secuencia de paridad $\{b_i\}_{i=1}^{n-k}$. Por definición, tenemos

$$\frac{B(D)}{M(D)} = \frac{1 + D + D^2 + D^3}{1 + D + D^3}$$

Al multiplicar en forma cruzada, obtenemos

$$(1 + D + D^2 + D^3)M(D) = (1 + D + D^3)B(D)$$

lo cual, bajo la inversión en el dominio del tiempo, produce

$$m_i + m_{i-1} + m_{i-2} + m_{i-3} + b_i + b_{i-1} + b_{i-3} = 0 \tag{10.70}$$

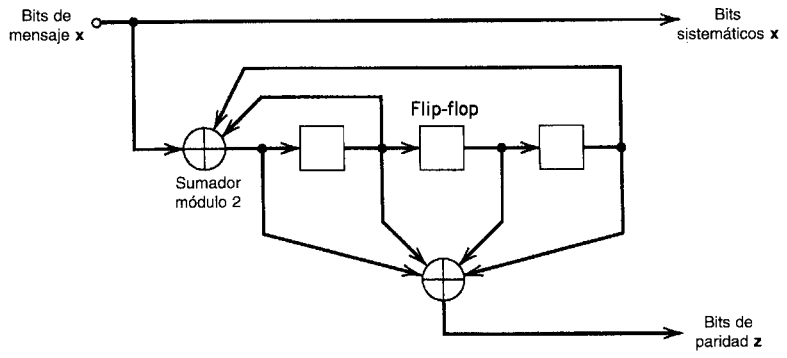


FIGURA 10.26 Ejemplo del codificador convolucional sistemático recursivo (RSC) de ocho estados.

donde la adición es módulo 2. La expresión (10.70) es la ecuación de verificación de paridad, la cual satisface al codificador convolucional de la figura 10.26 en cada etapa de tiempo i. ◀

En la figura 10.25, la secuencia de datos de entrada se aplica directamente al codificador 1, y la versión reordenada pseudoaleatoria de la misma secuencia de datos se aplica al codificador 2. Los bits sistemáticos (es decir, los bits del mensaje original) y los dos conjuntos de bits verificadores de paridad generados por los dos codificadores constituyen la salida del codificador turbo. Si bien los códigos constituyentes son convolucionales, en realidad los turbo son códigos de bloque con el tamaño de bloque determinándose a partir del tamaño del intercalador. Además, puesto que ambos codificadores RSC en la figura 10.25 son lineales, es posible describir los códigos turbo como *códigos de bloque lineales*.

La naturaleza de bloque de los códigos turbo plantea una cuestión práctica: ¿cómo conocer el principio y el final de una palabra de código? La práctica común consiste en asignar al codificador los valores iniciales correspondientes al *estado de puros ceros* y luego codificar los datos. Después de codificar cierto número de bits de datos se agregan varios bits de cola, de manera que se consiga que el codificador regrese al estado de puros ceros al final de cada bloque; de ahí en adelante se repite el ciclo. Los métodos de terminación de los códigos turbo incluyen lo siguiente:

- Un método simple consiste en terminar el primer código RSC en el codificador y dejar el segundo indeterminado. Una desventaja de este método es que los bits al final del bloque debidos al segundo código RSC son más vulnerables al ruido que el resto de ellos. El trabajo experimental ha demostrado que los códigos turbo presentan una nivelación en el desempeño cuando aumenta la SNR. Este comportamiento no es similar a un nivel de errores, sino que tiene la apariencia de un nivel de errores comparado con una disminución escalonada en el desempeño ante los errores a una SNR baja. Este *nivel de errores* resulta afectado por varios factores, siendo el dominante la elección del intercalador.
- Un método más refinado[14] implica terminar de manera simétrica ambos códigos constituyentes en el codificador. Mediante el uso combinado de un buen intercalador y de una terminación dual, el nivel de errores puede reducirse en un orden de magnitud en comparación con el método de terminación simple.

En la versión original del codificador turbo, los bits de verificación de paridad generados por los dos codificadores en la figura 10.25 se eliminaron antes de la transmisión de datos por el canal para mantener la tasa a 1/2. Un *código con eliminación* se construye eliminando ciertos bits de verificación de paridad, incrementando de esa manera la tasa de datos. La eliminación es lo inverso de la extensión de un código. Sin embargo, debe subrayarse que el uso de un mapa de eliminación no es un requerimiento necesario para la generación de códigos turbo.

Lo novedoso en el esquema de codificación en paralelo de la figura 10.25 radica en el uso de códigos convolucionales sistemáticos recursivos (RSC) y la introducción de un intercalador pseudoaleatorio entre los dos codificadores. De ese modo un código turbo aparece esencialmente *aleatorio* para el canal en virtud del intercalador pseudoaleatorio, aunque posee suficiente estructura para que la decodificación sea realizable físicamente. La teoría de la codificación afirma que un código elegido al azar es capaz de aproximarse a la capacidad de canal de Shannon, siempre que el tamaño del bloque sea suficientemente grande.[15] Ésta es en realidad la razón detrás del impresionante desempeño de los códigos turbo, según se explica a continuación.

■ DESEMPEÑO DE LOS CÓDIGOS TURBO

La figura 10.27 muestra el desempeño ante errores de un código turbo de tasa 1/2 con un gran tamaño de bloque correspondiente a la transmisión de datos binaria por un canal AWGN.[16] El

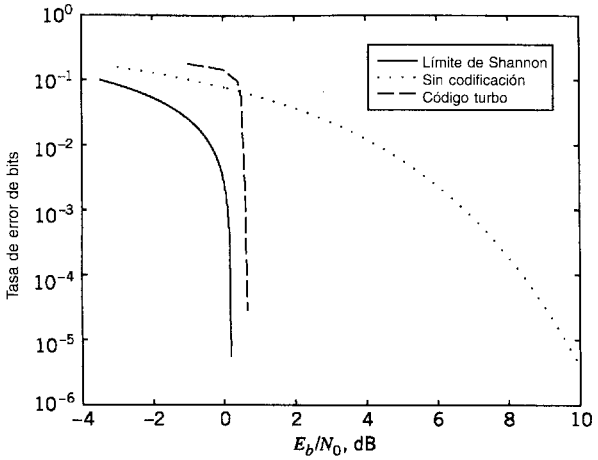


FIGURA 10.27 Desempeños ante el ruido de un código turbo de tasa 1/2 y transmisión sin codificación en un canal AWGN; la figura incluye también el límite teórico de Shannon de la capacidad del canal para la tasa de código $r = 1/2$.

código utiliza un intercalador de 65,536 de tamaño y un decodificador basado en BCJR; los detalles de este codificador se presentan más adelante en la sección. Se utilizaron dieciocho iteraciones de la decodificación turbo en el cálculo.

Con fines de comparación, la figura 10.27 incluye también otras dos curvas para el mismo canal AWGN:

- Transmisión no codificada (es decir, tasa de código $r = 1$).
- Límite teórico de Shannon para la tasa de código 1/2, el cual sigue de la figura 9.18*b*.

Con base en la figura 10.27, podemos sacar dos importantes conclusiones:

1. Si bien la tasa de error de bits para la transmisión de codificación turbo es bastante superior que para la transmisión sin codificación a baja E_b/N_0, la tasa de error de bits correspondiente a la transmisión con codificación turbo disminuye muy rápido una vez que se ha alcanzado el valor crítico de E_b/N_0.
2. A una tasa de error de bits de 10^{-5}, el código turbo es menor que 0.5 dB con respecto al límite teórico de Shannon.

Sin embargo, advierta que alcanzar este desempeño sumamente impresionante requiere que sea grande el tamaño del intercalador o, equivalentemente, la longitud del bloque del código turbo. Además, el gran número de iteraciones que se necesita para mejorar el desempeño incrementa la latencia del decodificador. Esta desventaja se debe a que el procesamiento digital de la información no conduce de inmediato por sí solo a la aplicación de realimentación, lo cual es un rasgo distintivo del decodificador turbo.

Ahora que tenemos una apreciación relativa al impresionante desempeño de los códigos turbo, el escenario está listo para una explicación de cómo se ejecuta realmente la decodificación turbo.

DECODIFICACIÓN TURBO

Los códigos turbo derivan su nombre distintivo de la analogía del algoritmo de decodificación con el principio del "motor turbo". La figura 10.28*a* muestra la estructura básica del decodificador turbo. Éste opera sobre versiones ruidosas de los bits sistemáticos y los dos conjuntos de bits de verificación de paridad en las dos etapas de decodificación para producir una estimación de los bits del mensaje originales.

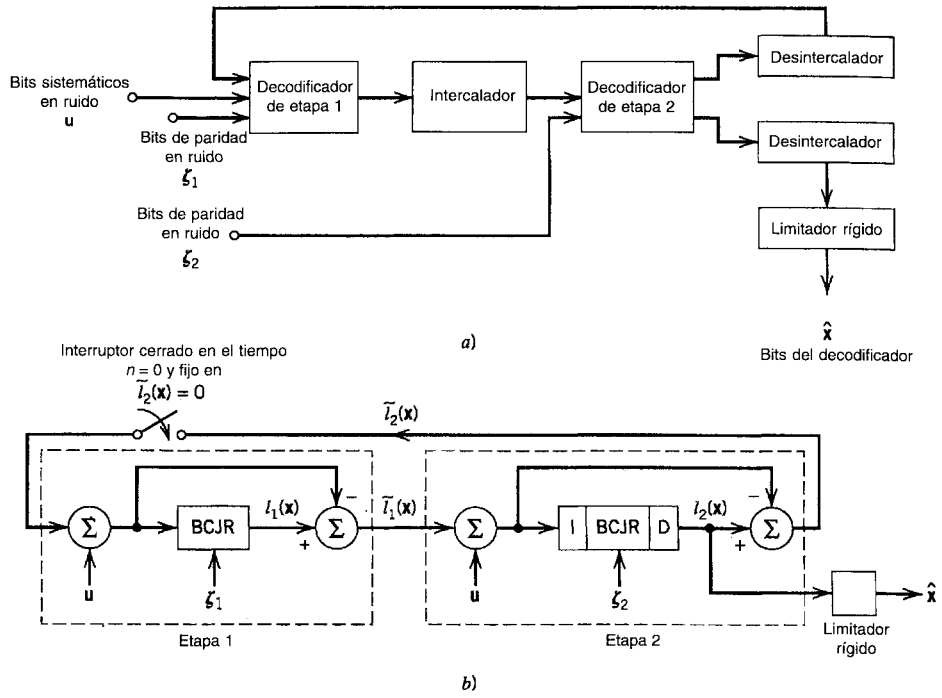


FIGURA 10.28 *a*) Diagrama de bloques de un decodificador turbo. *b*) Forma extrínseca del decodificador turbo, donde I representa al intercalador, D al desintercalador y BCJR corresponde al algoritmo BCJR para la decodificación MAP logarítmica.

Cada una de las dos etapas de decodificación utiliza un *algoritmo BCJR*,[17] el cual fue inventado originalmente por Bahl, Cocke, Jelinek y Raviv (de ahí el nombre) para resolver un problema de *detección de probabilidad a posteriori máxima (MAP)*. El algoritmo BCJR difiere del de Viterbi en dos aspectos fundamentales:

1. El BCJR es un algoritmo de decodificación de *entrada flexible-salida flexible*, con una recursión simple hacia adelante que incluye decisiones flexibles; la recursión finaliza con una decisión rígida mediante la cual se retiene una trayectoria sobreviviente particular de entre varias. En términos de cómputo, el algoritmo BCJR es entonces más complejo que el de Viterbi debido a la recursión hacia atrás.
2. El algoritmo BCJR es un decodificador MAP en el sentido de que minimiza los errores de bit al estimar las probabilidades *a posteriori* de los bits individuales en una palabra de código; para construir la secuencia de datos original, las salidas flexibles del algoritmo BCJR son de limitación rígida. Por otra parte, el algoritmo de Viterbi es un estimador de secuencia de máxima verosimilitud en cuanto a que maximiza la función de verosimilitud para la secuencia completa, no cada bit. Como tal, la tasa promedio de errores de bit del algoritmo BCJR puede ser un poco mejor que el algoritmo de Viterbi; en ningún caso es peor.

Lo más importante es que la formulación del algoritmo BCJR se sustenta en las suposiciones fundamentales de que 1) la codificación del canal, a saber, la codificación convolucional efectuada en el transmisor, se modela como un *proceso de Markov*, y 2) el canal es sin memoria. En el contexto de nuestra discusión presente, la suposición markoviana quiere decir que si un código

puede representarse como un enramado, entonces el estado presente del enramado depende únicamente del estado pasado y del bit de entrada. (Un tratamiento matemático del algoritmo BCJR se presenta después en esta sección.)

Antes de describir la operación del decodificador turbo de dos etapas de la figura 10.28*a*, encontramos deseable introducir la noción de información extrínseca. La representación más conveniente para este concepto es una razón de verosimilitud logarítmica, en cuyo caso la información extrínseca se calcula como la diferencia entre dos cocientes de verosimilitud logarítmica, como se describe en la figura 10.29. Formalmente, la *información extrínseca* generada por una etapa de decodificación para un conjunto de bits sistemáticos (mensaje), se define como la diferencia entre la razón de verosimilitud logarítmica calculada a la salida de esa etapa de decodificación y la *información intrínseca* representada por una proporción de verosimilitud logarítmica realimentada a la entrada de la etapa de decodificación. En efecto, la información extrínseca constituye la información incremental que se gana al explotar las dependencias que existen entre un bit de mensaje de interés y los bits de datos en bruto entrantes procesados por el decodificador.

Con esta base es posible describir el flujo de información en el codificador turbo de dos etapas de la figura 10.28*a* de una manera *extrínseca simétrica*, como se ilustra en la figura 10.28*b*. La primera etapa de decodificación utiliza el algoritmo BCJR para producir una estimación flexible del bit sistemático x_j, expresado como la proporción de verosimilitud logarítmica

$$l_1(x_j) = \log\left(\frac{P(x_j = 1|\mathbf{u}, \zeta_1, \widetilde{l}_2(\mathbf{x}))}{P(x_j = 0|\mathbf{u}, \zeta_1, \widetilde{l}_2(\mathbf{x}))}\right), \qquad j = 1, 2, \ldots, k \tag{10.71}$$

donde $\mathbf{u}$ es el conjunto de bits sistemáticos ruidosos, ζ_1 es el conjunto de bits de verificación de paridad ruidosos generados por el codificador 1, y $\widetilde{l}_2(\mathbf{x})$ es la información extrínseca en torno al conjunto de bits de mensaje $\mathbf{x}$ derivados de la segunda etapa de decodificación y realimentados a la primera. Suponiendo que los k bits de mensaje son estadísticamente independientes, la proporción de verosimilitud logarítmica total a la salida de la primera etapa de decodificación es entonces

$$l_1(\mathbf{x}) = \sum_{j=1}^{k} l_1(x_j) \tag{10.72}$$

Por tanto, la información extrínseca en torno a los bits de mensaje derivados de la primera etapa de decodificación es

$$\widetilde{l}_1(\mathbf{x}) = l_1(\mathbf{x}) - \widetilde{l}_2(\mathbf{x}) \tag{10.73}$$

donde $\widetilde{l}_2(\mathbf{x})$ va a definirse.

Antes de la aplicación en la segunda etapa de decodificación, la información extrínseca $\widetilde{l}_1(\mathbf{x})$ se reordena para compensar el intercalamiento pseudoaleatorio introducido en el codificador turbo. Además, los bits de verificación de paridad ruidosos ζ_2 generados por el codificador 2 se usan como entrada. Por tanto, al usar el algoritmo BCJR, la segunda etapa de decodificación produce una estimación flexible más refinada de los bits de mensaje $\mathbf{x}$. Dicha estimación se reintercala para producir la proporción de verosimilitud logarítmica total $l_2(\mathbf{x})$. La información

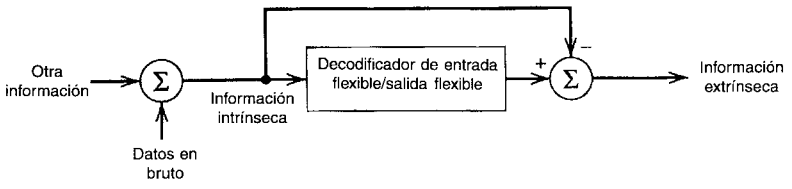


FIGURA 10.29 Ilustración del concepto de información extrínseca.

extrínseca $\tilde{l}_2(\mathbf{x})$ que se realimenta a la primera etapa de decodificación es entonces

$$\tilde{l}_2(\mathbf{x}) = l_2(\mathbf{x}) - \tilde{l}_1(\mathbf{x}) \tag{10.74}$$

donde $\tilde{l}_1(\mathbf{x})$ es a su vez definida por la ecuación (10.73), y $l_2(\mathbf{x})$ es la proporción de verosimilitud logarítmica calculada mediante la segunda etapa. Específicamente, para el j-ésimo elemento del vector $\mathbf{x}$, tenemos

$$l_2(x_j) = \log_2\left(\frac{P(x_j = 1 \mid \mathbf{u}, \zeta_2, \tilde{l}_1(\mathbf{x}))}{P(x_j = 0 \mid \mathbf{u}, \zeta_2, \tilde{l}_1(\mathbf{x}))}\right), \quad j = 1, 2, \ldots, k \tag{10.75}$$

Mediante la aplicación $\tilde{l}_2(\mathbf{x})$ en la primera etapa, el lazo de realimentación alrededor del par de etapas de decodificación está por eso cerrado. Note que si bien el conjunto de bits sistemáticos ruidosos $\mathbf{u}$ se aplica en realidad sólo en la primera etapa de decodificación, como en la figura 10.28*a*, al formular el flujo de información de la manera extrínseca simétrica que se describe en la figura 10.28*b*, encontramos que $\mathbf{u}$, en efecto, se aplica también en la segunda etapa de decodificación.

Una estimación de los bits de mensaje $\mathbf{x}$ se calcula al limitar en forma rígida la proporción de verosimilitud logarítmica $l_2(\mathbf{x})$ a la salida de la segunda etapa, como se indica mediante

$$\hat{\mathbf{x}} = \text{sgn}(l_2(\mathbf{x})) \tag{10.76}$$

donde la función signo opera individualmente en cada elemento de $l_2(\mathbf{x})$.

Para iniciar el algoritmo de decodificación turbo, simplemente dejamos $\tilde{l}_2(\mathbf{x}) = 0$ en la primera iteración del algoritmo; vea la figura 10.28*b*.

La motivación para alimentar únicamente información extrínseca de una etapa a la siguiente en el decodificador turbo de la figura 10.8, reside en mantener la mayor independencia estadística posible entre los bits de una iteración a la siguiente. La estrategia de decodificación de realimentación descrita aquí se fundamenta implícitamente en esta suposición. Si esta suposición de independencia estadística es estrictamente cierta, puede demostrarse que la estimación $\hat{\mathbf{x}}$ definida en la ecuación (10.76) tiende a la solución MAP cuando el número de iteraciones se aproxima a infinito.[18] La suposición de independencia estadística parece ser muy cercana a la verdad en la mayoría de casos encontrados en la práctica.

■ EL ALGORITMO BCJR

Para que una discusión de la decodificación turbo sea completa, conviene una exposición matemática del algoritmo BCJR para la estimación MAP.

Sea $x(t)$ la entrada a un codificador de enramado en el tiempo t. Sea $y(t)$ la salida correspondiente observada en el receptor. Advierta que $y(t)$ quizá incluya más de una observación; por ejemplo, un código de tasa $1/n$ produce n bits por cada bit de entrada, en cuyo caso tenemos un vector de observación n-dimensional. Considere que el vector de observación se denota por medio de

$$\mathbf{y}_{(1,t)} = [y(1), y(2), \ldots, y(t)]$$

Sea $\lambda_m(t)$ la probabilidad de que un estado $s(t)$ del codificador de enramado sea igual a m, donde $m = 1, 2, \ldots, M$. Es posible escribir entonces

$$\boldsymbol{\lambda}(t) = P[\mathbf{s}(t) | \mathbf{y}] \tag{10.77}$$

donde $\mathbf{s}(t)$ y $\boldsymbol{\lambda}(t)$ son vectores de M por 1. En ese caso, para un código convolucional lineal de

tasa $1/n$ con realimentación, como en el código RSC, la probabilidad de que el símbolo "1" sea el bit de mensaje está dada por

$$P(x(t)=1\,|\,\mathbf{y})=\sum_{s\in\mathscr{T}_A}\lambda_s(t) \tag{10.78}$$

donde $\mathscr{T}_A$ es el conjunto de transiciones que corresponde al símbolo "1" a la entrada, y $\lambda_s(t)$ es la componente s de $\boldsymbol{\lambda}(t)$.

Defina la *estimación directa* de las probabilidades de estado como el vector de M por 1

$$\boldsymbol{\alpha}(t)=P(\mathbf{s}(t)|\,\mathbf{y}_{(1,t)}) \tag{10.79}$$

donde el vector de observación $\mathbf{y}_{(1,t)}$ se define antes. También defina la *estimación hacia atrás* de las probabilidades de estado como el vector M por 1

$$\boldsymbol{\beta}(t)=P(\mathbf{s}(t)\,|\,\mathbf{y}_{(t,k)}) \tag{10.80}$$

donde

$$\mathbf{y}_{(t,k)}=[y(t),\,y(t+1),\,\ldots,\,y(k)]$$

Los vectores $\boldsymbol{\alpha}(t)$ y $\boldsymbol{\beta}(t)$ son estimaciones de las probabilidades de estado en el tiempo t basadas en los datos pasados y futuros, respectivamente. Es posible entonces formular el *teorema de separabilidad* como sigue:

> Las probabilidades de estado en el tiempo t se relacionan con el estimador hacia adelante $\boldsymbol{\alpha}(t)$ y con el estimador hacia atrás $\boldsymbol{\beta}(t)$ mediante el vector
>
> $$\boldsymbol{\lambda}(t)=\frac{\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)}{\|\boldsymbol{\alpha}(t)\cdot\boldsymbol{\alpha}(t)\|_1} \tag{10.81}$$
>
> donde $\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)$ es el vector producto de $\boldsymbol{\alpha}(t)$ y $\boldsymbol{\beta}(t)$, y $\|\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)\|_1$ es la norma L_1 de este producto vectorial.

El *producto vectorial* $\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)$ (no debe confundirse con el producto interior) se define en términos de los elementos individuales de $\boldsymbol{\alpha}(t)$ y $\boldsymbol{\beta}(t)$ mediante

$$\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)=\begin{bmatrix}\alpha_1(t)\beta_1(t)\\ \alpha_2(t)\beta_2(t)\\ \vdots\\ \alpha_M(t)\beta_M(t)\end{bmatrix} \tag{10.82}$$

y la *norma* L_1 de $\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)$ se define mediante

$$\|\boldsymbol{\alpha}(t)\cdot\boldsymbol{\beta}(t)\|_1=\sum_{m=1}^{M}\alpha_m(t)\beta_m(t) \tag{10.83}$$

El teorema de separabilidad afirma que la distribución de estados en el tiempo t dado el pasado es independiente de la distribución de estados en el tiempo t dado el futuro, lo cual se satisface de manera intuitiva recordando la suposición markoviana para la codificación del canal, la cual es fundamental para el algoritmo BCJR. Asimismo, este teorema proporciona la base de una forma simple de combinar las estimaciones hacia adelante y hacia atrás con el fin de obtener una descripción completa de las probabilidades de estado.

Para continuar, dejamos que la probabilidad de la transición de estado en el tiempo t se defina por medio de

$$\gamma_{m',m}(t)=P(s(t)=m,\mathbf{y}(t)\,|\,s(t-1)=m') \tag{10.84}$$

y denotamos la matriz de probabilidades de transición de M por M como

$$\boldsymbol{\Gamma}(t) = \{\gamma_{m',m}(t)\} \tag{10.85}$$

Podemos formular entonces el *teorema de recursión* como sigue:

La estimación hacia adelante $\boldsymbol{\alpha}(t)$ y la estimación hacia atrás $\boldsymbol{\beta}(t)$ se calculan recursivamente como

$$\boldsymbol{\alpha}^T(t) = \frac{\boldsymbol{\alpha}^T(t-1)\boldsymbol{\Gamma}(t)}{\|\boldsymbol{\alpha}^T(t-1)\boldsymbol{\Gamma}(t)\|_1} \tag{10.86}$$

y

$$\boldsymbol{\beta}(t) = \frac{\boldsymbol{\Gamma}(t+1)\boldsymbol{\beta}(t+1)}{\|\boldsymbol{\Gamma}(t+1)\boldsymbol{\beta}(t+1)\|_1} \tag{10.87}$$

donde el exponente T denota transposición de matrices.

Los teoremas de separabilidad y recursión en conjunto definen el algoritmo BCJR para el cálculo de probabilidades *a posteriori* de los estados y transiciones de un enramado de código, dado el vector de observación. Utilizando estas estimaciones, las proporciones de probabilidad necesarias para la decodificación turbo pueden calcularse entonces efectuando las sumatorias sobre subconjuntos seleccionados de estados según se requiera.

10.9 *Experimento de computadora: decodificación turbo*

Dos propiedades constituyen el sello distintivo de los códigos turbo:

Propiedad 1:

El desempeño ante los errores del decodificador turbo mejora con el número de iteraciones del algoritmo de decodificación. Esto se consigue alimentando información extrínseca de la salida de la primera etapa de decodificación en la entrada de la segunda etapa de decodificación en la trayectoria hacia adelante y alimentando información extrínseca de la salida de la segunda etapa en la entrada de la primera etapa en la trayectoria hacia atrás, y permitiendo después que el proceso de decodificación iterativa tome su curso natural en respuesta al mensaje ruidoso y a los bits de paridad recibidos.

Propiedad 2

El decodificador turbo es capaz de aproximarse al límite teórico de Shannon de la capacidad del canal de una manera factible computacionalmente; esta propiedad se ha demostrado experimentalmente, pero no teóricamente.

La propiedad 2 requiere que la longitud del bloque de un código turbo sea grande. Desafortunadamente, una demostración de esta propiedad requiere el uso de puestas en práctica complejas del algoritmo de decodificación turbo que están más allá del objetivo de este libro. Por tanto, centraremos nuestra atención en la demostración de la propiedad 1 en este experimento de computadora.

De ese modo, como el objetivo primario de este experimento de computadora, deseamos utilizar la puesta en práctica MAP logarítmica del algoritmo BCJR para demostrar la propiedad

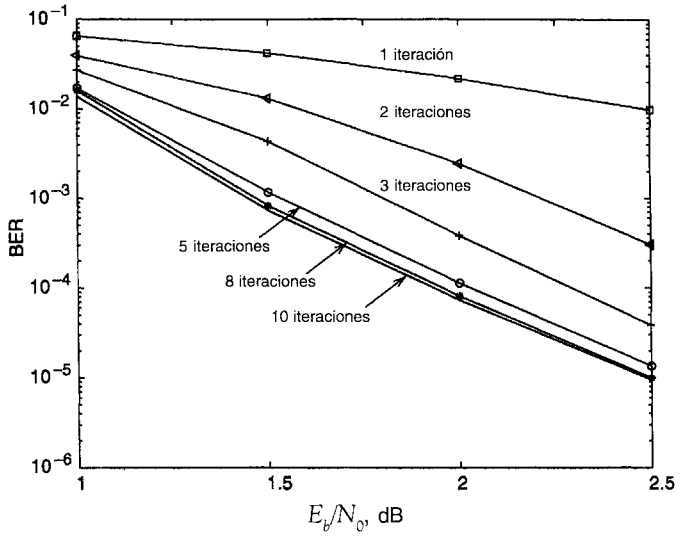


FIGURA 10.30 Resultados del experimento de computadora para la decodificación turbo, relativo a un número creciente de iteraciones.

1 de la decodificación turbo. El único deterioro del canal supuesto en el experimento es el ruido blanco gaussiano aditivo. Los detalles del codificador y el decodificador turbo son como sigue:

Codificador turbo (descrito en la figura 10.25):

Codificador 1: codificador convolucional [1, 1, 1]
Codificador 2: codificador convolucional [1, 0, 1]
Longitud de bloque (es decir, intercalador): 1200 bits

Decodificador turbo (descrito en la figura 10.28):

El algoritmo BCJR para la decodificación MAP logarítmica.

El experimento se ejecutó para $E_b/N_0 = 1, 1.5, 2$ y 2.5 dB, con número variable de iteraciones para cada E_b/N_0. En cada ensayo del experimento se calculó el número de errores de bit después de acumular un total de 20 bloques de datos (cada uno de 1200 bits de largo) que habían sido corrompidos por ruido. La probabilidad de error se evaluó después de eso como la razón entre los errores de bit y el número total de bits codificados. Advierta que en este cálculo, muchos de los bloques de bits codificados se decodificaron de modo correcto.

Los resultados del experimento se grafican en la figura 10.30. Las siguientes observaciones son posibles a partir de dicha figura:

1. Para una E_b/N_0 dada, la probabilidad de error disminuye con el número creciente de iteraciones, lo que confirma la propiedad 1 de la decodificación turbo.
2. Después de ocho iteraciones, no hay mejora importante en el desempeño de la decodificación.
3. Para un número fijo de iteraciones, la probabilidad de error disminuye con E_b/N_0 creciente, lo cual era lo esperado.

10.10 *Códigos de verificación de paridad de baja densidad*[19]

Los códigos turbo, explicados en la sección 10.8, y los códigos de verificación de paridad de baja densidad (LDPC), analizados en esta sección, pertenecen a una amplia familia de técnicas de

codificación de control de errores conocida como *códigos compuestos*. Las dos ventajas más importantes de los códigos LDPC con respecto a los códigos turbo son:

- Ausencia de palabras de código de bajo peso.
- Decodificación iterativa de complejidad inferior.

En lo que respecta al tema de las palabras de código de bajo peso, usualmente encontramos que un número pequeño de palabras de código en un código turbo están indeseablemente cercanas a la palabra de código determinada. Debido a esta proximidad en los pesos, de vez en cuando el ruido del canal provoca que la palabra de código transmitida sea equivocada en relación con una palabra de código cercana. En realidad, este comportamiento es el responsable del nivel de error (por lo común alrededor de una tasa de error de bits de 10^{-5} a 10^{-6}) que se mencionó antes. En contraste, los códigos LDPC pueden construirse con facilidad de manera que éstos no tengan tales palabras de código de *peso bajo* y sea posible entonces que alcancen tasas de error de bits *sumamente pequeñas*. El problema del nivel de error en los códigos turbo puede resolverse mediante un diseño cuidadoso del intercalador.

Volviendo al tema de la complejidad de la decodificación, observamos que la complejidad computacional de un decodificador turbo está dominada por el algoritmo BCJR, el cual opera sobre el enramado en el código convolucional que se utiliza en el codificador. El número de cálculos en cada recursión del algoritmo BCJR se escala linealmente con el número de estados en el enramado. Los códigos turbo que se usan comúnmente emplean enramados con 16 o más estados. En contraste, los códigos LDPC utilizan un enramado de verificación de paridad simple que tiene sólo dos estados. Por ello los decodificadores para los códigos LDPC son bastante más simples que los correspondientes a los decodificadores turbo. Además, al ser posible ponerlos en paralelo, la codificación LDPC puede realizarse a velocidades superiores que la decodificación turbo.

Sin embargo, una objeción práctica al uso de los códigos LDPC es que en el caso de grandes longitudes de bloque, su complejidad de codificación es elevada en comparación con los códigos turbo.

■ CONSTRUCCIÓN DE CÓDIGOS LDPC

Los códigos LDPC se especifican mediante una matriz de verificación de paridad denotada por $\mathbf{A}$, la cual es *dispersa*; es decir, consta principalmente de ceros y de un pequeño número de unos. En particular, hablamos de códigos LDPC (n, t_c, t_r), donde n denota la longitud de bloque, t_c el peso (es decir, el número de unos) en cada columna de la matriz $\mathbf{A}$, y t_r el peso de cada renglón con $t_r > t_c$. La tasa de un código LDPC de estas características es

$$r = 1 - \frac{t_c}{t_r} \tag{10.88}$$

cuya validez es posible justificar de la manera siguiente. Dejemos que ρ denote la *densidad* de unos en la matriz de verificación de paridad $\mathbf{A}$. En ese caso, siguiendo la terminología que se presentó en la sección 10.3, es posible fijar

$$t_c = \rho(n - k)$$

y

$$t_r = \rho n$$

donde $(n - k)$ es el número de renglones en $\mathbf{A}$ y n es el número de columnas (es decir, la longitud del bloque). Por tanto, al dividir t_c entre t_r, obtenemos

$$\frac{t_c}{t_r} = 1 - \frac{k}{n}$$

Por definición, la tasa de código de un código de bloque es k/n, de lo cual surge el resultado de la ecuación (10.88). Sin embargo, para que se cumpla este resultado, los renglones de $\mathbf{A}$ deben ser *linealmente independientes*.

Las *gráficas bipartidas* representan de manera adecuada la estructura de los códigos LDPC. En la figura 10.31 se muestra una de tales gráficas para el código de ejemplo de $n = 10$, $t_c = 3$ y $t_r = 5$. Los nodos del lado izquierdo en la gráfica de la figura 10.31 son *variables*, lo cual corresponde a elementos de la palabra de código. Los del lado derecho de la gráfica son *nodos de verificación*, que corresponden a un conjunto de restricciones de verificación de paridad que satisfacen las palabras de código en el código. Los códigos LDPC del tipo ejemplificado por la gráfica de la figura 10.31 se dice que serán *uniformes* en cuanto a que todos los nodos de tipo similar tienen exactamente el mismo grado. En la gráfica de ejemplo de la figura 10.31, el grado de los nodos variables es $t_c = 3$ y el de los nodos de verificación corresponde a $t_r = 5$. Cuando la longitud de bloque n se acerca a infinito, cada nodo de verificación se conecta a una fracción sumamente pequeña de los nodos variables, de lo que se desprende el término *baja densidad*.

La matriz $\mathbf{A}$ se construye colocando al azar unos en $\mathbf{A}$, considerando las *restricciones de regularidad*:

- Cada columna contiene un pequeño número fijo, t_c, de unos.
- Cada renglón contiene un pequeño número fijo, t_r, de unos.

En la práctica, estas restricciones de regularidad a menudo se violan ligeramente con el fin de evitar renglones linealmente dependientes en la matriz de verificación de paridad $\mathbf{A}$.

A diferencia de los códigos de bloque lineales que se presentaron en la sección 10.3, la matriz de verificación de paridad $\mathbf{A}$ de los códigos LDPC *no* es sistemática (es decir, no cuenta con bits de verificación de paridad que aparecen en forma diagonal), de ahí el uso de un símbolo diferente del que se utilizó en la sección 10.3. A pesar de eso, para los fines de codificación, es posible obtener una matriz generadora $\mathbf{G}$ para los códigos LDPC mediante *eliminación gaussiana* realizada en la aritmética módulo 2; este procedimiento se ilustra después en el ejemplo 10.9. Siguiendo la terminología que se presentó en la sección 10.3, un vector de código $\mathbf{c}$ de 1 por n se divide primero como

$$\mathbf{c} = [\mathbf{b} \vdots \mathbf{m}]$$

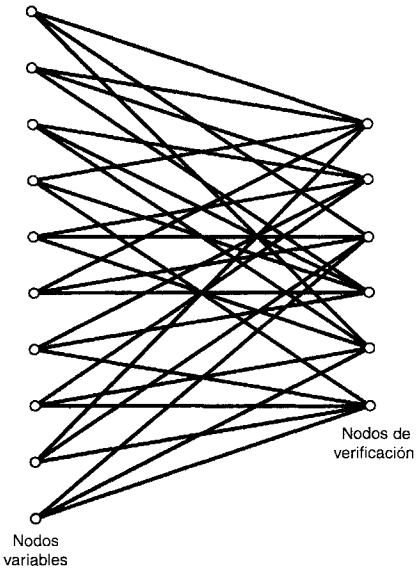


FIGURA 10.31 Gráfica bipartida del código LDPC (10, 3, 5).

donde **m** es el vector de mensaje de k por 1 y **b** es el vector de paridad de $(n - k)$ por 1; vea la ecuación (10.9). De manera correspondiente, la matriz de verificación de paridad **A** se divide como

$$\mathbf{A}^T = \begin{bmatrix} \mathbf{A}_1 \\ \cdots\cdots \\ \mathbf{A}_2 \end{bmatrix} \tag{10.89}$$

donde $\mathbf{A}_1$ es una matriz cuadrada de dimensiones $(n - k) \times (n - k)$ y $\mathbf{A}_2$ es una matriz rectangular de dimensiones $k \times (n - k)$; la transposición simbolizada mediante el exponente T se utiliza en la división de la matriz **A** por conveniencia en la presentación. Al imponer la restricción de la ecuación (10.16) sobre el código LDPC es posible escribir

$$[\mathbf{b} \vdots \mathbf{m}] \begin{bmatrix} \mathbf{A}_1 \\ \cdots\cdots \\ \mathbf{A}_2 \end{bmatrix} = 0$$

o, equivalentemente,

$$\mathbf{b}\mathbf{A}_1 + \mathbf{m}\mathbf{A}_2 = 0 \tag{10.90}$$

Recuerde de la ecuación (10.7) que los vectores **m** y **b** están relacionados por

$$\mathbf{b} = \mathbf{m}\mathbf{P}$$

donde **P** es la matriz de coeficientes. Por tanto, sustituyendo esta relación en la ecuación (10.90), encontramos de inmediato que, para cualquier vector de mensaje **m** distinto de 0, la matriz de coeficientes de los códigos LDPC satisface la condición

$$\mathbf{P}\mathbf{A}_1 + \mathbf{A}_2 = 0$$

la cual se cumple para *todos* los vectores de mensaje distintos de cero y, en particular, para **m** en la forma $[0 \cdots 0\ 1\ 0 \cdots 0]$, que aislará los renglones individuales de la matriz generadora.

Resolviendo esta ecuación para la matriz **P**, obtenemos

$$\mathbf{P} = \mathbf{A}_2\mathbf{A}_1^{-1} \tag{10.91}$$

donde $\mathbf{A}_1^{-1}$ es la *inversa* de la matriz $\mathbf{A}_1$, la cual se define de manera natural en la aritmética módulo 2. Por último, la matriz generadora de los códigos LDPC está definida por

$$\begin{aligned} \mathbf{G} &= [\mathbf{P} \vdots \mathbf{I}_k] \\ &= [\mathbf{A}_2\mathbf{A}_1^{-1} \vdots \mathbf{I}_k] \end{aligned} \tag{10.92}$$

donde $\mathbf{I}_k$ es la matriz identidad de k por k; vea la ecuación (10.12).

Es importante advertir que si consideramos la matriz de verificación de paridad **A** para algún código LDPC arbitrario y sólo escogemos $(n - k)$ columnas de **A** al azar para formar una matriz cuadrada $\mathbf{A}_1$, *no* hay garantía de que $\mathbf{A}_1$ será *no singular* (es decir, existirá la inversa $\mathbf{A}^{-1}$), incluso si los renglones de **A** son linealmente independientes. De hecho, para un código LDPC representativo con gran longitud de bloque n, tal $\mathbf{A}_1$ elegida al azar es altamente improbable que sea no singular, debido a que es sumamente probable que al menos un renglón de $\mathbf{A}_1$ será de puros ceros. Desde luego, cuando los renglones de **A** son linealmente independientes, existirá *algún* conjunto de $(n - k)$ columnas **A** que formarán una $\mathbf{A}_1$ no singular, como se ilustra en el ejemplo 10.9. En algunos métodos de construcción de códigos LDPC es posible que se garantice que las primeras $(n - k)$ columnas de **A** produzcan una $\mathbf{A}_1$ no singular o que al menos así suceda con alta probabilidad, aunque esto *no* es cierto en general.

EJEMPLO 10.9 Código LDPC (10, 3, 5)

Considere la gráfica bipartida de la figura 10.31 perteneciente a un código LDPC (10, 3, 5). La matriz de verificación de paridad del código está definida por

$$\mathbf{A} = \underbrace{\left[\begin{array}{cccccc} 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right.}_{\mathbf{A}_1^T} \underbrace{\left.\begin{array}{cccc} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{array}\right]}_{\mathbf{A}_2^T}$$

la cual parece ser aleatoria, mientras se mantengan las restricciones de regularidad: $t_c = 3$ y $t_r = 5$. Dividiendo la matriz $\mathbf{A}$ de la manera descrita en la ecuación (10.89):

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Para obtener la inversa de la matriz $\mathbf{A}_1$, utilizamos primero la ecuación (10.90) para escribir

$$\underbrace{[b_0, b_1, b_2, b_3, b_4, b_5]}_{\mathbf{b}} \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{A}_1} = \underbrace{[u_0, u_1, u_2, u_3, u_4, u_5]}_{\mathbf{u} = \mathbf{m}\mathbf{A}_2}$$

donde hemos introducido el vector $\mathbf{u}$ para denotar el producto de matrices $\mathbf{m}\mathbf{A}_2$. Utilizando eliminación gaussiana, la matriz $\mathbf{A}_1$ se transforma en una *forma diagonal inferior* (es decir, todos los elementos sobre la diagonal principal son cero), según indica

$$\mathbf{A}_1 \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Esta transformación se consigue mediante las siguientes sumas módulo 2 efectuadas sobre las columnas de la matriz cuadrada $\mathbf{A}_1$:

- Las columnas 1 y 2 se suman a la columna 3.
- La columna 2 se suma a la columna 4.
- Las columnas 1 y 4 se suman a la columna 5.
- Las columnas 1, 2 y 5 se suman a la columna 6.

De manera correspondiente, el vector $\mathbf{u}$ se transforma como

$$\mathbf{u} \rightarrow [u_0, u_1, u_0 + u_1 + u_2, u_1 + u_3, u_0 + u_3 + u_4, u_0 + u_1 + u_4 + u_5]$$

Por tanto, premultiplicando la matriz transformada $\mathbf{A}_1$ por el vector de paridad $\mathbf{b}$, utilizando eliminaciones sucesivas en la aritmética de módulo 2 operando en sentido inverso, y poniendo las soluciones para los elementos del vector de paridad $\mathbf{b}$ en términos de los elementos del vector $\mathbf{u}$ en forma de matriz, obtenemos

$$\underbrace{[u_0, u_1, u_2, u_3, u_4, u_5]}_{\mathbf{u}} \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}}_{\mathbf{A}_1^{-1}} = \underbrace{[b_0, b_1, b_2, b_3, b_4, b_5]}_{\mathbf{b}}$$

La inversa de la matriz $\mathbf{A}_1$ es entonces

$$\mathbf{A}_1^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

El producto de matrices $\mathbf{A}_2\mathbf{A}_1^{-1}$ es (utilizando el valor dado de $\mathbf{A}_2$ y el valor de $\mathbf{A}_1^{-1}$ que acaba de encontrarse)

$$\mathbf{A}_2\mathbf{A}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Por último, empleando la ecuación (10.92), el generador del código LDPC (10, 3, 5) es

$$\mathbf{G} = \underbrace{\left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{array}\right.}_{\mathbf{A}_2\mathbf{A}_1^{-1}} \underbrace{\left.\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]}_{\mathbf{I}_k}$$

Es importante admitir que el código LDPC descrito en este ejemplo persigue únicamente ilustrar el procedimiento implicado en la generación de un código de este tipo. En la práctica, la longitud de bloque n es varias órdenes de magnitud mayor que la que se considera en este ejemplo. Además, en la construcción de la matriz $\mathbf{A}$ es posible que restrinjamos a todos los pares de columnas

para que tengan un *traslape de matriz* (es decir, producto interno de cualesquiera dos columnas en una matriz **A**) que no exceda 1; una restricción de este tipo, por encima de las restricciones de regularidad, se espera que mejore el desempeño de los códigos LDPC. Infortunadamente, con una longitud de bloque pequeña como la que se consideró en este ejemplo, resulta difícil satisfacer este requisito adicional. ◀

Distancia mínima de los códigos LDPC

En la práctica, la longitud de bloque de un código LDPC es grande, variando de 10^3 a 10^6, lo que significa que el número de palabras de código en un código particular es correspondientemente elevado. En consecuencia, el análisis algebraico de los códigos LDPC es bastante difícil. Resulta mucho más productivo efectuar un *análisis estadístico* sobre un conjunto de códigos LDPC. Dicho análisis nos permite realizar enunciados estadísticos acerca de ciertas propiedades de los códigos miembro en el conjunto. Además, un código LDPC con estas propiedades puede encontrarse con alta probabilidad mediante una selección aleatoria a partir del conjunto.

Entre estas propiedades, la distancia mínima de los códigos miembro es de particular interés. De la sección 10.3 recordamos que la distancia mínima de un código de bloque lineal es, por definición, la distancia de Hamming más pequeña entre cualquier par de vectores de código en el código. Sobre un conjunto de códigos LDPC, la distancia mínima de un código miembro es naturalmente una variable aleatoria. En otra parte[20] se demuestra que cuando aumenta la longitud de bloque n, para $t_c \geq 3$ y $t_r > t_c$ fijas, la distribución de probabilidad de la distancia mínima puede sobreacotarse mediante una función que se aproxima a la función escalón unitario a una fracción fija $\Delta_{t_c t_r}$ de la longitud de bloque n. Por tanto, para n grande, prácticamente todos los códigos LDPC en el conjunto tienen una distancia mínima de al menos $n\Delta_{t_c t_r}$. La tabla 10.10 presenta la tasa r y la $\Delta_{t_c t_r}$ de códigos LDPC para diferentes valores del par de peso (t_c, t_r). A partir de esta tabla vemos que para $t_c = 3$ y $t_r = 6$, la tasa de código r alcanza su valor más alto de 1/2 y la fracción $\Delta_{t_c t_r}$ llega a su valor más pequeño, y de ahí la elección preferida de $t_c = 3$ y $t_r = 6$ en el diseño de códigos LDPC.

Decodificación probabilística de códigos LDPC

En el transmisor, un vector de mensaje **m** se codifica en un vector de código $\mathbf{c} = \mathbf{mG}$, donde **G** es la matriz generadora para un par de peso (t_c, t_r) especificado y, en consecuencia, distancia mínima $d_{\text{mín}}$. El vector **c** se transmite por un canal ruidoso para producir el vector recibido.

$$\mathbf{r} = \mathbf{c} + \mathbf{e}$$

donde **e** es el vector de error debido al ruido del canal; vea la ecuación (10.17). Por construcción, la matriz **A** es una matriz de paridad del código LDPC; es decir, $\mathbf{AG}^T = \mathbf{0}$. Dado el vector

Tabla 10.10[a] *La tasa r y el término fraccionario $\Delta_{t_c t_r}$ de los códigos LDPC para los pesos variables t_c y t_r*

t_c	t_r	*Tasa r*	$\Delta_{t_c t_r}$
5	6	0.167	0.255
4	5	0.2	0.210
3	4	0.25	0.122
4	6	0.333	0.129
3	5	0.4	0.044
3	6	0.5	0.023

[a]Adaptado de Gallager [256] con autorización del IEEE.

recibido **r**, el problema de decodificación bit por bit radica en determinar el vector más probable $\hat{\mathbf{c}}$ que satisface la condición $\hat{\mathbf{c}}\mathbf{A}^T = \mathbf{0}$.

En lo que sigue, un bit se refiere a un elemento del vector recibido **r**, y una verificación corresponde a un renglón de una matriz **A**. Sea $\mathcal{J}(i)$ el conjunto de bits que participa en la verificación i. Sea $\mathcal{J}(i)$ el conjunto de verificaciones en las cuales participa el bit j. Un conjunto $\mathcal{J}(i)$ que excluye al bit j se denota mediante $\mathcal{J}(i)\backslash j$. De igual modo, un conjunto $\mathcal{J}(j)$ que excluye la verificación i se denota por medio de $\mathcal{J}(j)\backslash i$.

El algoritmo de decodificación tiene dos pasos alternativos: paso horizontal y paso vertical, los cuales se ejecutan a lo largo de los renglones y columnas de la matriz **A**, respectivamente. En el curso de estos pasos, se actualizan de manera alternativa dos cantidades probabilísticas asociadas con elementos distintos de cero de la matriz **A**. Una cantidad, denotada mediante P_{ij}^x, define la probabilidad de que el bit j sea el símbolo x (es decir, símbolo 0 o 1), dada la información obtenida mediante las verificaciones realizadas en el paso horizontal, con excepción de la verificación i. La segunda cantidad, denotada por Q_{ij}^x, define la probabilidad de que se satisfaga la verificación i, dado que el bit j se fija en el valor x y los otros bits tienen las probabilidades $P_{ij'}: j' \in \mathcal{J}(i)\backslash j$.

El algoritmo de decodificación LDPC se realiza entonces de la manera siguiente:[21]

Asignación de valores iniciales

Las variables P_{ij}^0 y P_{ij}^1 se igualan a las probabilidades a priori P_j^0 y P_j^1 de los símbolos 0 y 1, respectivamente, con $P_j^0 + P_j^1 = 1$.

Paso horizontal

En el paso horizontal del algoritmo, ejecutamos a través de las verificaciones i. Defina

$$\Delta P_{ij} = P_{ij}^0 - P_{ij}^1$$

Para cada par de peso (i, j), calcule

$$\Delta Q_{ij} = \prod_{j' \in \mathcal{J}(i)\backslash j} \Delta P_{ij'}$$

Por tanto, establezca

$$Q_{ij}^0 = \frac{1}{2}(1 + \Delta Q_{ij})$$

$$Q_{ij}^1 = \frac{1}{2}(1 - \Delta Q_{ij})$$

Paso vertical

En el paso vertical del algoritmo, los valores de las probabilidades P_{ij}^0 y P_{ij}^1 se actualizan utilizando las cantidades calculadas en el paso horizontal. En particular, para cada bit j calcule

$$P_{ij}^0 = \alpha_{ij} p_j^0 \prod_{i' \in \mathcal{J}(j)\backslash i} Q_{i'j}^0$$

$$P_{ij}^1 = \alpha_{ij} p_j^1 \prod_{i' \in \mathcal{J}(j)\backslash i} Q_{i'j}^1$$

donde el factor de escalamiento α_{ij} se elige para hacer

$$P_{ij}^0 + P_{ij}^1 = 1$$

En el paso vertical, es posible también actualizar las *probabilidades pseudoposteriores*:

$$P_j^0 = \alpha_j p_j^0 \prod_{i \in \mathcal{J}(j)} Q_{ij}^0$$

$$P_j^1 = \alpha_j p_j^1 \prod_{i \in \mathcal{J}(j)} Q_{ij}^1$$

donde α_j se elige para hacer

$$P_j^0 + P_j^1 = 1$$

Las cantidades obtenidas en el paso vertical se utilizan para calcular una estimación tentativa $\hat{\mathbf{c}}$. Si la condición $\hat{\mathbf{c}}\mathbf{A}^T = \mathbf{0}$ se cumple, se termina el algoritmo de decodificación. En otro caso, el algoritmo vuelve al paso horizontal. Si después de algún número máximo de iteraciones (por ejemplo 100 o 200) no hay decodificación válida, se declara una falla de la decodificación. El procedimiento de decodificación descrito aquí constituye un caso especial del *algoritmo de suma-producto general de baja complejidad*.

Establecido en términos simples, el algoritmo de suma-producto deja pasar cantidades probabilísticas entre los nodos de verificación y los nodos de variable de la gráfica bipartida. En virtud de que cada restricción de verificación de paridad puede representarse mediante un codificador convolucional simple con un bit de memoria, encontramos que los decodificadores LDPC son más simples de poner en práctica que los decodificadores turbo, como se estableció antes.

Sin embargo, en términos de desempeño es posible afirmar lo siguiente a la luz de los resultados experimentales informados en la literatura: los códigos LDPC uniformes no parecen acercarse al límite de Shannon como lo hacen sus contrapartes códigos turbo.

10.11 Códigos irregulares

Los códigos turbo que se explicaron en la sección 10.8 y los códigos LDPC presentados en la sección 10.10 son códigos regulares, cada uno en su propia forma individual. El desempeño de corrección de errores de ambos códigos por un canal ruidoso puede mejorarse de manera sustancial utilizando sus formas irregulares respectivas.

En un código turbo estándar con su codificador, como se muestra en la figura 10.25, el intercalador asigna cada bit sistemático a un bit de entrada único del codificador convolucional 2. En contraste, los *códigos turbo irregulares*[22] utilizan un diseño especial de intercalador que asigna algunos bits sistemáticos a bits de entrada múltiples del codificador convolucional. Por ejemplo, cada 10% de los bits sistemáticos es posible que se asigne a ocho entradas del codificador

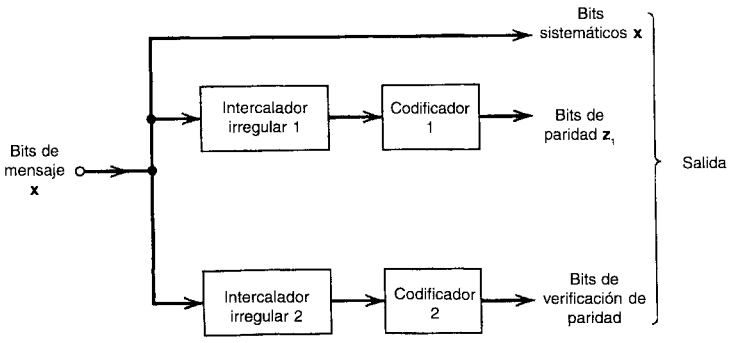


FIGURA 10.32 Diagrama de bloques de un codificador turbo irregular.

convolucional en lugar de una sola. Como se muestra en la figura 10.32, se utilizan *intercaladores irregulares* similares en ambas trayectorias de codificación convolucional para generar los bits de verificación de paridad $\mathbf{z}_1$ y $\mathbf{z}_2$ en respuesta a los bits de mensaje $\mathbf{x}$. Los códigos turbo irregulares se decodifican de modo similar a la de los códigos turbo regulares.

Para construir un *código LDPC irregular*,[23] los grados de los nodos variable y de verificación en la gráfica bipartida se seleccionan de acuerdo con alguna distribución. Por ejemplo, podemos tener un código LDPC irregular con la siguiente representación gráfica:

- Una mitad de los nodos variables tiene grado 5 y la otra mitad grado 3.
- Una mitad de los nodos de verificación tiene grado 6 y la otra mitad grado 8.

Para una longitud de bloque determinada y una secuencia de grado dada, definimos un conjunto de códigos eligiendo los bordes (es decir, las conexiones entre los nodos variables y de verificación) de manera aleatoria. Específicamente, los bordes que emanan de los nodos variables se enumeran en algún orden arbitrario, y lo mismo ocurre para los bordes que emanan de los nodos de verificación.

En la figura 10.33 se grafica el desempeño con respecto a los errores de los siguientes códigos:[24]

- Código LDPC irregular: $k = 50{,}000$, $n = 100{,}000$, tasa $= 1/2$
- Código turbo (regular): $k = 65{,}536$, $n = 131{,}072$, y tasa $= 1/2$
- Código turbo irregular: $k = 65{,}536$, $n = 131{,}072$, y tasa $= 1/2$

donde k es el número de bits de mensaje y n es la longitud del bloque. Los polinomios generadores correspondientes a los dos codificadores convolucionales en los códigos turbo regular/irregular son como sigue:

Codificador 1: $g(D) = 1 + D^4$

Codificador 2: $g(D) = 1 + D + D^2 + D^3 + D^4$

La figura 10.33 incluye también el límite teórico correspondiente sobre la capacidad del canal para la tasa de código $r = 1/2$.

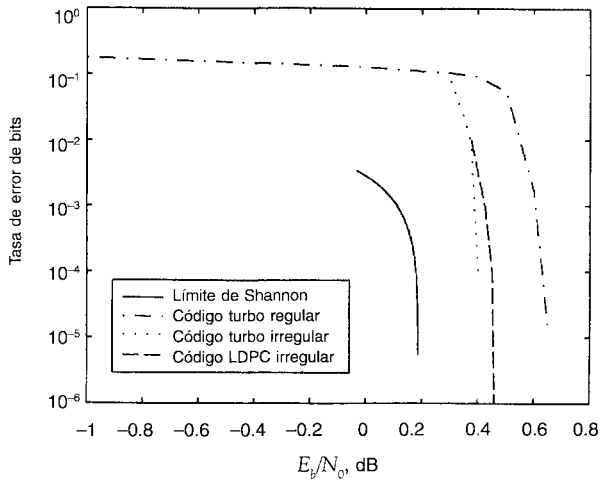


FIGURA 10.33 Desempeño ante el ruido del código turbo regular, el código turbo irregular y el código irregular de verificación de paridad y baja densidad (LDPC), en comparación con el límite de Shannon para la tasa de código $r = 1/2$.

Con base en los resultados presentados en la figura 10.33, son posibles las siguientes observaciones:

- El código LDPC irregular supera el desempeño del código turbo regular porque se acerca más al límite teórico de Shannon en 0.175 dB.
- Entre los tres códigos presentados ahí, el código turbo irregular es el mejor con respecto a que sólo se encuentra a 0.213 dB del límite teórico de Shannon.

10.12 *Resumen y discusión*

En este capítulo estudiamos las técnicas de codificación de control de errores que por sí solas se han establecido como herramientas indispensables en la comunicación digital confiable por canales ruidosos. El efecto de los errores que ocurren durante la transmisión se reduce agregando redundancia a los datos antes de la transmisión de una manera controlada. La redundancia se utiliza para permitir a un decodificador en el receptor detectar y corregir errores.

Las técnicas de codificación de control de errores pueden dividirse en dos familias ampliamente definidas:

1. *Códigos algebraicos*, los cuales se sustentan en la estructura algebraica abstracta incorporada en el diseño de los códigos para la decodificación en el receptor. Los códigos algebraicos incluyen a los de Hamming, a los de longitud máxima, a los BCH y a los Reed-Solomon. Estos códigos particulares comparten dos propiedades:

   Propiedad de linealidad, la suma de cualesquiera dos palabras de código en el código es también una palabra de código.

   Propiedad cíclica, cualquier corrimiento cíclico de una palabra de código también es una palabra de código en el código.

   Los códigos de Reed-Solomon son muy poderosos, así como capaces de combatir tanto los errores aleatorios como los de ráfaga; encuentran aplicaciones en ambientes difíciles, tales como las comunicaciones espaciales y los discos compactos.
2. *Códigos probabilísticos*, los cuales se fundamentan en métodos probabilísticos para su decodificación en el receptor. Los códigos probabilísticos incluyen a los de enramado, a los turbo y a los de verificación de paridad de baja densidad. En particular, la decodificación se basa en uno u otro de los dos métodos básicos, como se resume a continuación:

   Entrada flexible-salida rígida, la cual se ejemplifica mediante el algoritmo de Viterbi que lleva a cabo la estimación de la secuencia de máxima verosimilitud en la codificación de los códigos de enramado.

   Entrada flexible-salida flexible, que se ejemplifica mediante el algoritmo BCJR que realiza la estimación *a posteriori* máxima sobre una base de bit por bit en la decodificación de los códigos turbo, o una forma especial del algoritmo de suma-producto en la decodificación de los códigos de verificación de paridad de baja densidad.

Los códigos de enramado combinan la codificación convolucional lineal y la modulación para permitir ganancias de codificación significativas con respecto a la modulación multinivel convencional no codificada sin sacrificar eficiencia del ancho de banda. Los códigos turbo y los códigos de verificación de paridad de baja densidad comparten las siguientes propiedades:

- Codificación aleatoria de un tipo de bloque lineal.
- Desempeño ante los errores exactamente dentro del límite teórico de Shannon con respecto a la capacidad del canal de una manera físicamente realizable.

En términos prácticos, los códigos turbo y los códigos de verificación de paridad de baja densidad, han posibilitado conseguir ganancias de codificación del orden de 10 dB, las cuales resultaban antes inalcanzables. Estas ganancias de codificación pueden explotarse hasta extender drásticamente la

gama de receptores de comunicación digital, aumentar en forma sustancial las tasas de bit de los sistemas de comunicación digital o disminuir de manera significativa la energía de la señal transmitida por símbolo. Los beneficios anteriores tienen implicaciones importantes en el diseño de comunicaciones inalámbricas y en comunicaciones espaciales, sólo por mencionar dos aplicaciones importantes de las comunicaciones digitales. En realidad, los códigos turbo ya se han estandarizado para aplicarse en enlaces de comunicación espaciales y en sistemas de comunicación inalámbricos.

NOTAS Y BIBLIOGRAFÍA

1. En lo que se refiere a una discusión introductoria de la corrección de errores mediante codificación, vea el capítulo 2 de Lucky [96]; consulte también el libro de Adamék [3] y el artículo de Bhargava [206]. El libro clásico sobre la codificación de control de errores es el de Peterson y Weldon [118]. La codificación de control de errores se analiza también en el libro clásico de Gallager [46]. Los libros de Lin y Costello [91], Micheleson y Levesque [109], MacWilliams y Sloane [98] y Wilson se dedican también a la codificación del control de errores. Para una colección de artículos fundamentales sobre el desarrollo de la teoría de codificación, vea el libro editado por Berlekamp [17].
2. En relación con un estudio de diversos esquemas ARQ, vea Lin, Costello y Miller [306].
3. En medicina, el término *síndrome* se utiliza para describir un patrón de síntomas que ayuda en el diagnóstico de una enfermedad. En la codificación, el patrón de errores desempeña el papel de la enfermedad y la falla en la verificación de paridad el del síntoma. Este uso de *síndrome* fue acuñado por Hagelbarger [271].
4. Los primeros códigos de corrección de errores (conocidos como códigos Hamming) fueron inventados por Hamming casi al mismo tiempo que la concepción de la teoría de información de Shannon; para mayores detalles, vea el artículo clásico de Hamming [274].
5. En cuanto a una descripción de los códigos BCH y de sus algoritmos de decodificación, vea Lin y Costello ([91], pp. 141-183) y MacWilliams y Sloane ([98], pp. 257-293). La tabla 10.6 sobre códigos BCH binarios se adaptó de Lin y Costello [91].
6. Los códigos Reed–Solomon reciben ese nombre en honor a sus inventores: vea el artículo clásico de estos autores [349]. Para detalles de los códigos Reed-Solomon, consulte MacWilliams y Sloane ([98], pp. 294-306). El libro editado por Wicker y Bhargava [166] contiene una introducción a los códigos Reed-Solomon, un panorama histórico de estos códigos escrito por sus inventores, Irving S. Reed y Gustave Solomon, y las aplicaciones de estos códigos a la exploración del sistema solar, el disco compacto, los protocolos de solicitud de repetición automática y las comunicaciones de acceso múltiple de espectro disperso, así como capítulos sobre otros temas relacionados.
7. Los códigos convolucionales los presentó por primera vez, como una alternativa a los códigos de bloque, P. Elias [239].
8. El término *enramado* fue introducido por Forney [247].
9. En un artículo clásico, Viterbi [389] propuso un algoritmo de decodificación para los códigos convolucionales que se ha llegado a conocer como *algoritmo de Viterbi*. Éste fue reconocido por Forney [246, 247] como un decodificador de máxima verosimilitud. En Lin y Costello [91], Blahut [24] y Adamék [3] se presentan explicaciones accesibles del algoritmo de Viterbi.
10. Los códigos convolucionales catastróficos se explican en Benedetto, Biglieri y Castellani [12]. La tabla 10.8 se adaptó del libro de estos últimos.
11. En cuanto a mayores detalles de la evaluación de la ganancia de codificación asintótica para canales AWGN simétricos binarios y de entrada binaria, vea Viterbi y Omura ([162], pp. 242-252) y Lin y Costello ([91], pp. 322-329).
12. G. Ungerboeck inventó la modulación por codificación de enramado; la evolución histórica de esta misma se describe en Ungerboeck [384]. La tabla 10.9 se adaptó de este último artículo.

    La modulación por codificación de enramado puede considerarse como una forma de *codificación del espacio de la señal*, un punto de vista que se analiza a nivel introductorio en el capítulo 14 del

libro de Lee y Messerschmitt [85]. En lo que se refiere a un tratamiento amplio de la modulación por codificación de enramado, vea los libros de Biglieri, Divsalar, McLane y Simon [19] y Schlegel [132].

13. Los códigos turbo fueron creados por C. Berrou y A. Glavieux. El trabajo relativo a estos códigos lo motivaron dos artículos sobre códigos de corrección de errores: Battail [197] y Hagenauer y Hoeher [273]. La primera descripción de los códigos turbo utilizando argumentos heurísticos la presentaron Berrou, Glavieux y Thitmajshima [205] en un trabajo para una conferencia; vea también Berrou y Glavieux [203]. Para reflexiones sobre el trabajo previo en los códigos turbo y los desarrollos subsecuentes, consulte Berrou y Glavieux [204].

    En lo que se refiere a un libro sobre los fundamentos de los códigos turbo, vea Heegard y Wicker [64]. Utilizando un procedimiento reminiscente de la codificación aleatoria (vea la nota 15), Benedetto y Montorsi [200] han proporcionado explicaciones parciales del impresionante desempeño de los códigos turbo.

    En dos estudios independientes de los cuales se informa en los artículos de McEliece, MacKay y Cheng [321] y Kschischang y Frey [299], se demuestra que la decodificación turbo duplica un algoritmo de la inteligencia artificial debido a Pearl, que incluye la propagación de la creencia. El término *creencia* es otra manera de referirse a la probabilidad *a posteriori*. Estos dos artículos han abierto un nuevo camino de investigación, que vincula la decodificación turbo y las máquinas de aprendizaje. Para una explicación amplia de los códigos turbo vea el libro de Frey.

    Un intercalador pseudoaleatorio es básico para la operación de los códigos turbo. Daneshgaran y Mondin [228] presentan un procedimiento sistemático para diseñar intercaladores (es decir, permutadores) para los códigos turbo.

14. La terminación dual de los códigos turbo se analiza en Guinand y Lodge [270].
15. La codificación aleatoria se explica en Cover y Thomas [33], sección 8.7.
16. Las gráficas que se presentan en la figura 10.27 resultan de aquéllas en la figura 6.8 del libro de Frey.
17. A principios de la década de 1960, Baum y Welch obtuvieron un procedimiento iterativo para resolver el problema de estimación de parámetros, de lo cual surgió el nombre *algoritmo de Baum-Welch* (Baum y Petrie, Baum y otros). En el *algoritmo BCJR*, nombrado así en honor de Bahl, Cocke, Jelinek y Raviv [195], el algoritmo de Baum-Welch se aplica al problema de la decodificación de máxima verosimilitud de salida flexible de los códigos convolucionales.
18. La prueba de que la estimación $\hat{\mathbf{x}}$ en la ecuación (10.76) tiende a la solución MAP cuando el número de iteraciones tiende a infinito se explica en el artículo de Moher y Gulliver [324].
19. Los códigos de verificación de paridad de baja densidad (LDPC) los descubrió originalmente Gallager [47, 256]. Fueron redescubiertos de manera independiente por MacKay y Neal; vea también MacKay [316].

    En la década de 1960 y durante un buen tiempo a partir de ahí, las computadoras disponibles en esos tiempos no eran lo suficientemente poderosas para procesar las largas longitudes de bloque que resultaban necesarias para lograr un excelente desempeño con los códigos LDPC; de ahí la falta de interés en su uso por más de 20 años.
20. Para un tratamiento detallado del enunciado de que la distribución de probabilidad de la distancia mínima de un código LDPC tiende a la función escalón unitario de la longitud de bloque para ciertos valores del par de peso (t_c, t_r), vea Gallager [47, 256].
21. El algoritmo de decodificación de los códigos LDPC descritos aquí se obtuvo de MacKay y Neal [317].
22. Los códigos turbo irregulares fueron inventados por Frey y MacKay [251].
23. Los códigos LDPC irregulares fueron inventados por MacKay y otros [318] y Richardson y otros [355].
24. Los códigos cuyo desempeño se grafica en la figura 10.34, se deben a los siguientes creadores:
    - Códigos turbo regulares: Berrou y Glavieux [203]; Berrou y otros.
    - Códigos turbo irregulares: Frey y MacKay [251].
    - Códigos LDPC irregulares: Richardson y otros [355].

PROBLEMAS

Codificación de decisión flexible

10.1 Considere un canal sin memoria discreto de salida Q-aria y entrada binaria. Se dice que el canal será simétrico si la probabilidad de transición del canal $p(j|i)$ satisface la condición:

$$p(j|0) = p(Q-1-j|1), \quad j = 0, 1, \ldots, Q-1$$

Suponga que los símbolos 0 y 1 de entrada del canal son igualmente probables. Demuestre que los símbolos de salida del canal son también igualmente probables; es decir,

$$p(j) = \frac{1}{Q}, \quad j = 0, 1, \ldots, Q-1$$

10.2 Considere el demodulador cuantizado para las señales PSK binarias que se muestran en la figura 10.3*a*. El cuantizador es uno de cuatro niveles, normalizado como en la figura P10.2. Evalúe las probabilidades de transición del canal sin memoria discreto de salida cuaternaria y entrada binaria caracterizado de ese modo. Por tanto, demuestre que éste es un canal simétrico. Suponga que la energía de la señal transmitida por bit es E_b, y que el ruido blanco gaussiano aditivo tiene media cero y densidad espectral de ponencia $N_0/2$.

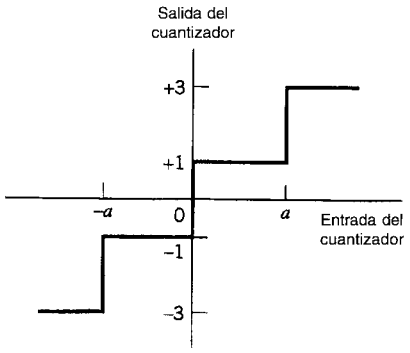


FIGURA P10.2

10.3 Considere un canal AWGN de entrada binaria, en el cual los símbolos binarios 1 y 0 son igualmente probables. Los símbolos binarios se transmiten por el canal mediante corrimiento de fase por llaveo. La energía del símbolo de código es E, y el AWGN tiene media cero y densidad espectral de potencia $N_0/2$. Demuestre que la probabilidad de transición del canal está dada por

$$p(y|0) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(y + \sqrt{\frac{2E}{N_0}}\right)^2\right], \quad -\infty < y < \infty$$

Códigos de bloques lineales y cíclicos

10.4 En un *código de verificación de paridad simple*, un único bit de paridad se agrega a un bloque de k bits de mensaje $(m_1, m_2, \ldots, m_k)$. El bit de paridad único b_1 se elige de manera que la palabra de código satisfaga la *regla de paridad par*:

$$m_1 + m_2 + \ldots + m_k + b_1 = 0, \quad \text{mod } 2$$

Para $k = 3$, establezca las 2^k palabras de código posibles en el código definido mediante esta regla.

10.5 Compare la matriz de verificación de paridad del código de Hamming (7, 4) considerado en el ejemplo 10.2 con la correspondiente a un código de repetición (4, 1).

10.6 Considere el código de Hamming (7, 4) del ejemplo 10.2. La matriz generadora **G** y la matriz de verificación de paridad **H** del código se describen en ese ejemplo. Demuestre que estas dos matrices satisfacen la condición

$$\mathbf{H}\mathbf{G}^T = 0$$

10.7 a) Para el código de Hamming (7, 4) descrito en el ejemplo 10.2, construya las ocho palabras de código en el código dual.

b) Determine la distancia mínima del código dual determinado en la parte *a*.

10.8 Considere el código de repetición (5, 1) del ejemplo 10.1. Evalúe el síndrome **s** para los siguientes patrones de errores:

a) La totalidad de los cinco posibles patrones de errores sencillos.

b) La totalidad de los 10 posibles patrones de errores dobles.

10.9 En una aplicación que requiere *sólo* detección de errores, es posible utilizar un código *no sistemático*. En este problema exploramos la generación de un código cíclico de estas características. Sea $g(X)$ el polinomio generador y $m(X)$ el polinomio del mensaje. Defina el polinomio del código $c(X)$ simplemente como

$$c(X) = m(X)g(X)$$

Por tanto, para un polinomio generador dado, es posible determinar de inmediato las palabras de código en el código. Para ilustrar este procedimiento, considere el polinomio generador para un código de Hamming (7, 4):

$$g(X) = 1 + X + X^3$$

Determine las 16 palabras de código en el código y confirme la naturaleza no sistemática de este mismo.

10.10 El polinomio $1 + X^7$ tiene $1 + X + X^3$ y $1 + X^2 + X^3$ como factores primitivos. En el ejemplo 10.3 utilizamos $1 + X + X^3$ como el polinomio generador para un código de Hamming (7, 4). En este problema consideramos la adopción de $1 + X^2 + X^3$ como el polinomio generador. Esto debe conducirnos al código de Hamming (7, 4) que es diferente del código que se analiza en el ejemplo 10.3. Formule el codificador y el calculador del síndrome para el polinomio generador:

$$g(X) = 1 + X^2 + X^3$$

Compare sus resultados con los del ejemplo 10.3.

10.11 Considere el código de Hamming (7, 4) definido por el polinomio generador:

$$g(X) = 1 + X + X^3$$

La palabra de código 0111001 se envía por un canal ruidoso, produciendo la palabra recibida 0101001 que tiene un solo error. Determine el polinomio del síndrome $s(X)$ para esta palabra recibida y demuestre que éste es idéntico al polinomio de error $e(X)$.

10.12 El polinomio generador de un código de Hamming (15, 11) está definido por:

$$g(X) = 1 + X + X^4$$

Formule el decodificador y el calculador del síndrome para este código, utilizando una forma sistemática para este último.

10.13 Considere el código de longitud máxima (15, 4) que es el doble del código de Hamming (15, 11) del problema 10.12. Realice lo siguiente:

a) Encuentre las conexiones de realimentación del codificador y compare sus resultados con los de la tabla 7.1 correspondientes a los códigos de longitud máxima que se presentan en el capítulo 7.

b) Determine el polinomio generador $g(X)$; por tanto, encuentre la secuencia de salida suponiendo el estado inicial 0001. Confirme la validez de su resultado reciclando el estado inicial a través del codificador.

10.14 Considere el código de Reed-Solomon (31, 15).
- **a)** ¿Cuántos bits hay en el símbolo del código?
- **b)** ¿Cuál es la longitud del bloque en bits?
- **c)** ¿Cuál es la distancia mínima del código?
- **d)** ¿Cuántos símbolos en el error puede corregir el código?

Códigos convolucionales

10.15 Un codificador convolucional tiene un registro de corrimiento sencillo con dos etapas (es decir, longitud de restricción $K = 3$), tres sumadores módulo 2 y un multiplexor de salida. Las secuencias del generador del codificador son las siguientes:

$$g^{(1)} = (1, 0, 1)$$
$$g^{(2)} = (1, 1, 0)$$
$$g^{(3)} = (1, 1, 1)$$

Dibuje el diagrama de bloques del codificador.
Nota: en los problemas del 10.16 al 10.23 se utiliza la misma secuencia de mensaje 10111... de modo que es posible comparar las salidas de los mismos codificadores para la misma entrada.

10.16 Considere el codificador convolucional de tasa $r = 1/2$ y longitud de restricción $K = 2$ de la figura P10.16. El código es sistemático. Determine la salida del codificador producida por la secuencia de mensaje 10111....

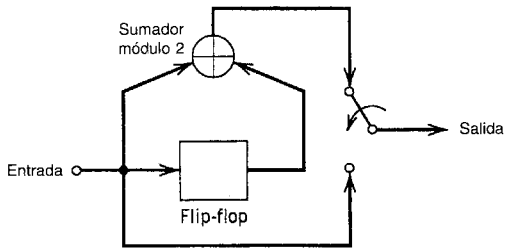


FIGURA P10.16

10.17 La figura P10.17 muestra el codificador para un código convolucional de tasa $r = 1/2$ y longitud de restricción $K = 4$. Determine la salida del codificador producida por la secuencia de mensaje 10111....

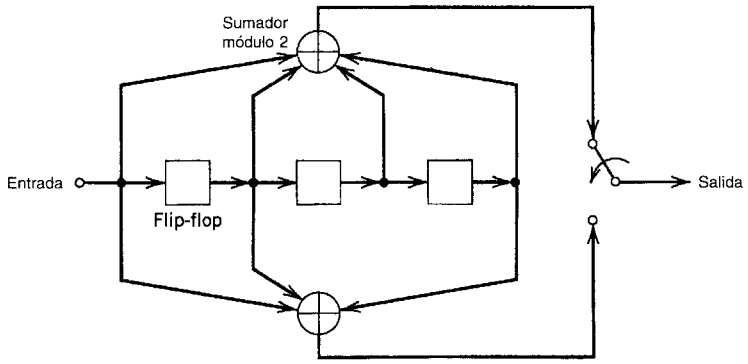


FIGURA P10.17

10.18 Considere el codificador de la figura 10.13*b* para un código convolucional de tasa $r = 2/3$ y longitud de restricción $K = 2$. Determine la secuencia de código producida por la secuencia de mensaje 10111....

10.19 Construya el árbol de código para el codificador convolucional de la figura P10.16. Trace la trayectoria a través del árbol que corresponda a la secuencia de mensaje 10111... y compare la salida del codificador con la que se determinó en el problema 10.16.

10.20 Construya el árbol de código para el codificador de la figura P10.17. Trace la trayectoria a través del árbol que corresponda a la secuencia de mensaje 10111.... Compare la salida resultante del codificador con la que se encontró en el problema 10.17.

10.21 Construya el diagrama de enramado para el codificador de la figura P10.17, suponiendo una secuencia de mensaje de longitud 5. Trace la trayectoria a través del enramado correspondiente a la secuencia de mensaje 10111.... Compare la salida resultante del codificador con la que se encontró en el problema 10.17.

10.22 Construya el diagrama de estado para el codificador de la figura P10.17. Empezando con el estado de puros ceros, trace la trayectoria que corresponde a la secuencia de mensaje 10111... y compare la secuencia de código resultante con la que se determinó en el problema 10.17.

10.23 Considere el codificador de la figura 10.13*b*.

a) Construya el diagrama de estado para este codificador.

b) Partiendo del estado de puros ceros, trace la trayectoria que corresponda a la secuencia de mensaje 10111.... Compare la secuencia resultante con la que se determinó en el problema 10.18.

10.24 Considerando el esquema de corrimiento mínimo por llaveo (MSK) como una máquina de estado finito, construya el diagrama de enramado para la MSK. (En el capítulo 6 se presenta una descripción de la MSK.)

10.25 El diagrama de enramado de un código convolucional de tasa 1/2 y longitud de restricción 3 se muestra en la figura P10.25. Se transmite la secuencia de puros ceros y la secuencia recibida corresponde a 100010000.... Utilizando el algoritmo de Viterbi, calcule la secuencia decodificada.

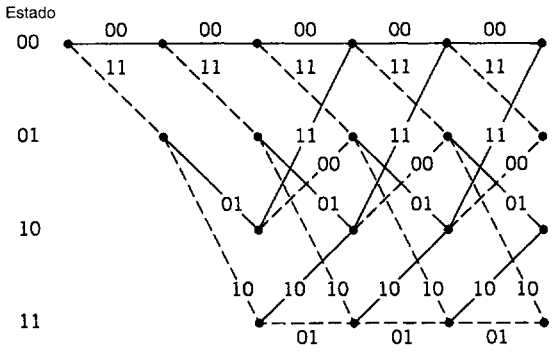


FIGURA P10.25

10.26 Considere un código convolucional de tasa 1/2 y longitud de restricción 7 con distancia libre $d_{\text{libre}} = 10$. Calcule la ganancia de codificación asintótica para los siguientes canales:

a) Canal simétrico binario.

b) Canal AWGN de entrada binaria.

10.27 En la sección 10.6 describimos el algoritmo de Viterbi para la decodificación de máxima verosimilitud de un código convolucional. Otra aplicación del algoritmo de Viterbi corresponde a la demodulación de máxima verosimilitud de una secuencia recibida que sufre la corrupción de la interferencia de intersímbolo debida a un canal disperso. La figura P10.27 muestra el diagrama de enramado para la interferencia de intersímbolo, suponiendo una secuencia de

datos binaria. El canal es discreto y lo describe la respuesta al impulso finita (1, 0.1). La secuencia recibida es (1.0, –0.3, –0.7, 0, ...). Emplee el algoritmo de Viterbi para determinar la versión decodificada de máxima verosimilitud de esta secuencia.

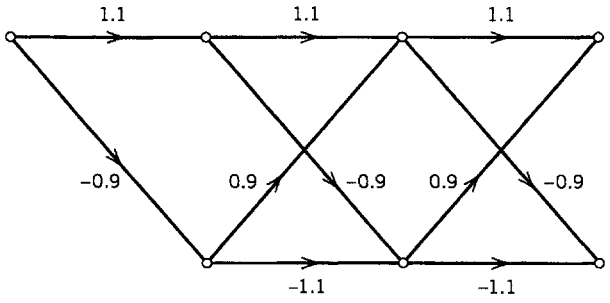


FIGURA P10.27

10.28 La figura P10.28 describe una constelación cruzada QAM-32. Ésta se divide en ocho subconjuntos. En cada etapa de la división, indique la distancia euclidiana dentro del subconjunto (la más corta).

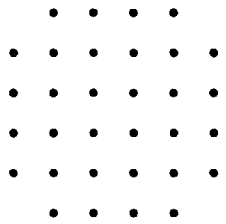


FIGURA P10.28

10.29 Como se explicó en la introducción de este capítulo, la codificación de canal puede utilizarse para reducir la E_b/N_0 que se requiere para un desempeño ante los errores preestablecidos o para reducir el tamaño de la antena receptora en relación con una E_b/N_0 prefijada. En este problema exploramos estos dos beneficios prácticos de la codificación al regresar al ejemplo 8.2 en el capítulo 8 acerca de los cálculos de potencia del enlace descendente de un sistema de comunicación de satélite local. En particular, suponemos que el diseño del enlace descendente incluye el uso de un esquema de codificación consistente en un codificador convolucional de tasa 1/2 con longitud $K = 7$ y decodificación de Viterbi. La ganancia de codificación de este esquema es igual a 5.1 dB, suponiendo el uso de cuantización flexible. De ese modo, realice lo siguiente:

a) Recalcule la proporción E_b/N_0 requerida del sistema.

b) Suponiendo que la proporción E_b/N_0 requerida permanece invariable, calcule la reducción en el tamaño de la antena de disco receptora que es posible gracias al uso de este esquema de codificación en el enlace descendente.

10.30 A diferencia de los códigos convolucionales considerados en este capítulo, recordamos del capítulo 6 que el código convolucional utilizado en el módem de banda de voz V.32 es *no lineal*. La figura P10.30 muestra el diagrama de circuito del codificador convolucional que se utiliza en este módem, el cual utiliza multiplicación módulo 2 y compuertas además de las sumas módulo 2 y los retardos. Explique la razón de la no linealidad del codificador de la figura P10.30, y utilice un ejemplo para ilustrar su explicación.

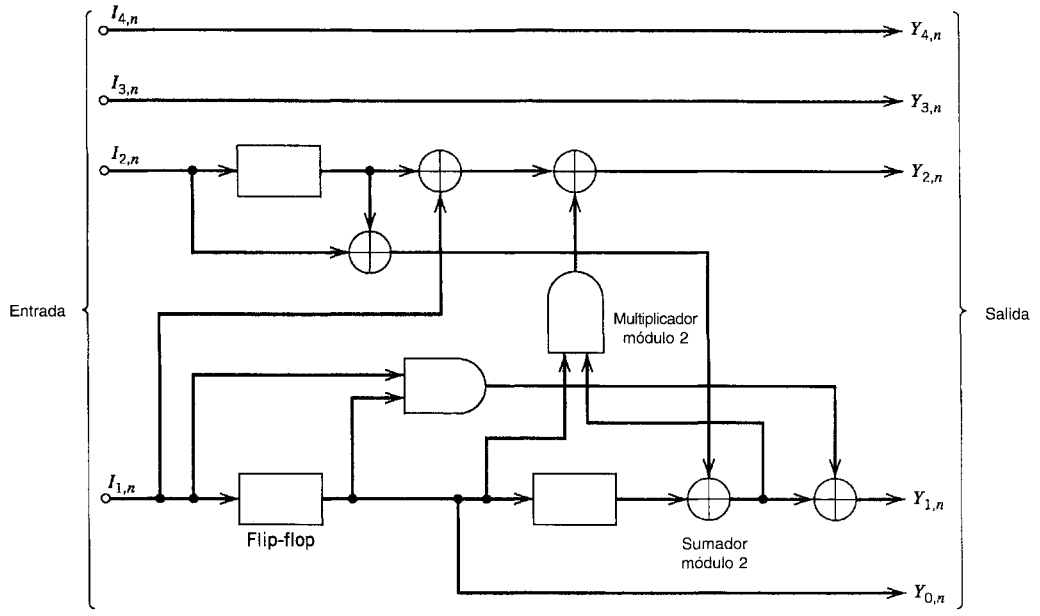


FIGURA P10.30

Códigos turbo

10.31 Sean $r_c^{(1)} = p/q_1$ y $r_c^{(2)} = p/q_2$ las tasas de código de los codificadores RSC 1 y 2 en el codificador turbo de la figura 10.25. Determine la tasa de código del codificador turbo.

10.32 La naturaleza de realimentación de los códigos constituyentes en el decodificador turbo de la figura 10.25 tiene la siguiente implicación: un error de bit sencillo corresponde a una secuencia infinita de errores del canal. Ilustre este fenómeno utilizando una secuencia de mensaje consistente en el símbolo 1 seguido por un número infinito de símbolos 0.

10.33 Considere las siguientes matrices generadoras para códigos turbo de tasa 1/2:

$$\text{codificador de 4 estados:} \quad g(D) = \left[1, \frac{1+D+D^2}{1+D^2}\right]$$

$$\text{codificador de 8 estados:} \quad g(D) = \left[1, \frac{1+D^2+D^3}{1+D+D^2+D^3}\right]$$

$$\text{codificador de 16 estados:} \quad g(D) = \left[1, \frac{1+D^4}{1+D+D^2+D^3+D^4}\right]$$

a) Construya el diagrama de bloques para cada uno de los codificadores RSC.

b) Establezca la ecuación de verificación de paridad asociada con cada codificador.

10.34 El codificador turbo de la figura 10.25 implica el uso de dos codificadores RSC.

a) Generalice este codificador para que abarque un total de M intercaladores.

b) Construya el diagrama de bloques del codificador turbo que aprovecha los M conjuntos de bits de verificación de paridad generados por tal generalización.

10.35 La decodificación turbo se sustenta en la realimentación de información extrínseca. El principio fundamental adherido al decodificador turbo radica en evitar la alimentación a una etapa de decodificación de información que provenga de la etapa misma. Explique la justificación de este principio en términos conceptuales.

10.36 Suponga que un receptor de comunicación consiste de dos componentes, un demodulador y un codificador. El demodulador se basa en un modelo de Markov del modulador y el canal combina-

dos; el decodificador se fundamenta en un modelo de Markov de un código de corrección de errores hacia adelante. Analice cómo puede aplicarse el principio turbo para construir un demodulador/decodificador conjunto para este sistema.

Experimento de computadora

10.37 En este experimento continuamos la investigación de los códigos turbo que se presentó en la sección 10.9 al evaluar el efecto del tamaño de bloque sobre el desempeño del decodificador ante el ruido.

Como antes, los dos codificadores convolucionales del codificador turbo son como sigue:

Codificador 1: [1, 1, 1]

Codificador 2: [1, 0, 1]

La E_b/N_0 transmitida es 1 dB. Se preestablece que los errores de bloque para la terminación no excederán de 15.

Con esta información previa, grafique la tasa de error de bits del decodificador turbo en función del número de iteraciones para dos tamaños de bloque (es decir, intercalador) diferentes: 200 y 400.

APÉNDICE 1

TEORÍA DE LA PROBABILIDAD

A1.1 *Conceptos probabilísticos*

La *teoría de la probabilidad* se fundamenta en fenómenos que, explícita o implícitamente, pueden modelarse mediante un experimento con un resultado que está sujeto a la *casualidad*. Además, si el experimento se repite, es posible que el resultado difiera debido a la influencia de un fenómeno aleatorio subyacente o del mecanismo de casualidad. Tal experimento se conoce como *experimento aleatorio*. Por ejemplo, el experimento puede ser la observación del resultado de lanzar al aire una moneda sin alterar. En este experimento, los resultados posibles del ensayo son "águila" o "sol".

Para ser más precisos en la descripción de un experimento aleatorio, se piden tres características:

1. El experimento es repetible bajo condiciones idénticas.
2. En un ensayo del experimento, el resultado es impredecible.
3. En un gran número de ensayos del experimento, los resultados exhiben *regularidad estadística*; es decir, se observa un patrón *promedio* definido de los resultados si el experimento se repite un gran número de veces.

MÉTODO DE FRECUENCIA RELATIVA

Considere que el *evento* A denota uno de los resultados posibles de un experimento aleatorio. Por ejemplo, en el experimento del lanzamiento de la moneda, el evento A puede representar "águila". Suponga que en n ensayos del experimento, el evento A ocurre $N_n(A)$ veces. Es posible entonces asignar el cociente $N_n(A)/n$ al evento A. Esta proporción recibe el nombre de *frecuencia relativa* del evento A. Claramente, la frecuencia relativa es un *número real no negativo menor que o igual a 1*; es decir,

$$0 \leq \frac{N_n(A)}{n} \leq 1 \tag{A1.1}$$

Si el evento A no ocurre en ninguno de los ensayos, $N_n(A)/n = 0$. Si, en cambio el evento A ocurre en todos los n ensayos, $N_n(A)/n = 1$.

Afirmamos que el experimento exhibe *regularidad estadística* si para *cualquier* secuencia de n ensayos la frecuencia relativa de $N_n(A)/n$ converge al mismo límite cuando n se vuelve grande. De esa manera parece natural para nosotros definir la *probabilidad del evento* A como

$$P(A) = \lim_{n \to \infty} \left(\frac{N_n(A)}{n} \right) \tag{A1.2}$$

El límite que se muestra en la ecuación (A1.2) no debe considerarse en un sentido matemático. En vez de eso, consideramos a la ecuación (A1.2) como el enunciado en que la probabilidad de un evento es la proporción a largo plazo de las veces que un evento particular A ocurre en una larga secuencia de ensayos. Por ejemplo, en el experimento del lanzamiento de la moneda, es posible esperar que de entre un millón de lanzamientos de una moneda sin alterar, alrededor de la mitad de ellos serán "águilas".

La probabilidad de un evento persigue representar la *posibilidad* de que un ensayo del experimento originará la ocurrencia de ese evento. En muchas aplicaciones de ingeniería y

juegos de azar, el uso de la ecuación (A1.2) para definir la probabilidad de un evento resulta aceptable. Sin embargo, en muchas otras aplicaciones, esta definición es inadecuada. Considere, por ejemplo, el análisis estadístico del mercado accionario: ¿cómo conseguiremos repetibilidad en un experimento de este tipo? Un método más satisfactorio consiste en establecer las propiedades que se espera que tenga cualquier medida de probabilidad, postulándolas como *axiomas* y utilizando después interpretaciones de frecuencia relativa para justificarlas.

■ AXIOMAS DE PROBABILIDAD

Cuando realizamos un experimento aleatorio, es natural para nosotros percatarnos de diversos resultados que es posible que se produzcan. En este contexto, es conveniente pensar un experimento y sus resultados posibles como la definición de un espacio y sus puntos. Con el k-ésimo resultado del experimento, digamos, asociamos un punto llamado el *punto de muestra*, que denotamos mediante s_k. La totalidad de los puntos de muestra correspondientes al agregado de todos los resultados posibles del experimento recibe el nombre del *espacio muestral*, el cual denotamos por medio de $\mathbf{S}$. Un evento corresponde ya sea a un punto muestral único o a un conjunto de puntos de la muestra. En particular, el espacio muestral completo $\mathbf{S}$ se denomina el *evento seguro*; el conjunto nulo $\varnothing$ recibe el nombre de *evento nulo* o *imposible*; y un punto de muestra único se conoce como un *evento elemental*.

Considere, por ejemplo, un experimento que implica el lanzamiento de un dado. En este experimento hay seis resultados posibles: la exhibición de uno, dos, tres, cuatro, cinco y seis puntos sobre la cara superior del dado. Asignando un punto de la muestra a cada uno de estos resultados posibles, tenemos un espacio muestral unidimensional que consiste en seis puntos muestrales, como se indica en la figura A1.1. El evento elemental que describe el enunciado "se exhibe un seis" corresponde al punto de muestra $\{6\}$. Por otra parte, el evento que describe el enunciado "se exhibe un número par de puntos" corresponde al conjunto $\{2, 4, 6\}$ del espacio muestral. Note que el término *evento* se usa de manera indistinta para describir el subconjunto o el enunciado.

Estamos listos para hacer una definición formal de probabilidad. Un *sistema de probabilidad* consta de los siguientes tres elementos:

1. *Un espacio muestral* $\mathbf{S}$ *de eventos elementales (resultados).*
2. *Una clase* $\mathscr{E}$ *de eventos que son subconjuntos de* $\mathbf{S}$.
3. *Una medida de probabilidad* $P(\cdot)$ *asignada en cada evento* A *en la clase* $\mathscr{E}$, *la cual tiene la siguientes propiedades:*

i) $$P(\mathbf{S}) = 1 \tag{A1.3}$$

ii) $$0 \leq P(A) \leq 1 \tag{A1.4}$$

iii) Si $A + B$ es la *unión de dos eventos mutuamente excluyentes* en la clase $\mathscr{E}$, entonces

$$P(A + B) = P(A) + P(B) \tag{A1.5}$$

Las propiedades anteriores se conocen como los *axiomas de probabilidad*. El axioma *i* establece que la probabilidad del evento seguro es la unidad. El axioma *ii* indica que la probabilidad de un

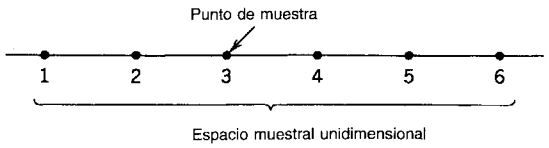


FIGURA A1.1 Espacio muestral para el experimento de lanzamiento de un dado.

evento es un número real no negativo que es menor o igual que la unidad. El axioma *iii* establece que la probabilidad de dos eventos mutuamente excluyentes es la suma de las probabilidades de los eventos individuales. Estos tres axiomas son suficientes para tratar con experimentos en que los espacios muestrales sean finitos.

Aun cuando el enfoque axiomático para la teoría de probabilidad es de naturaleza abstracta, los tres axiomas tienen sus propias interpretaciones de frecuencia relativa. El axioma *ii* corresponde a la ecuación (A1.1), y el axioma *i*, al caso límite de la ecuación (A1.1) cuando ocurre el evento A en todos los n ensayos. Para interpretar el axioma *iii*, advertimos que si el evento A ocurre $N_n(A)$ veces en n ensayos y el evento B ocurre $N_n(B)$ veces, entonces el evento unión "A o B" ocurre en $N_n(A) + N_n(B)$ ensayos (puesto que A y B nunca pueden ocurrir en el mismo ensayo). Por lo tanto, $N_n(A + B) = N_n(A) + N_n(B)$, y por ello tenemos

$$\frac{N_n(A+B)}{n} = \frac{N_n(A)}{n} + \frac{N_n(B)}{n}$$

que tiene una forma matemática similar a la del axioma *iii*.

Los axiomas *i*, *ii* y *iii* constituyen una definición implícita de probabilidad. Es posible utilizar estos axiomas para formular algunas otras propiedades básicas de la probabilidad, como se describe a continuación.

Propiedad 1

$$P(\overline{A}) = 1 - P(A) \tag{A1.6}$$

donde $\overline{A}$ *(denotando "no A") es el complemento del evento* A.

El uso de esta propiedad nos ayuda a investigar la *no ocurrencia de un evento*. Para probarlo, expresamos el espacio muestral S como la unión de dos eventos mutuamente excluyentes A y $\overline{A}$:

$$\mathsf{S} = A + \overline{A}$$

Entonces, el uso de los axiomas *i* y *iii* produce

$$1 = P(A) + P(\overline{A})$$

a partir de la cual se obtiene la ecuación (A1.6).

Propiedad 2

Si M *eventos mutuamente excluyentes* $A_1, A_2, \ldots, A_M$ *tienen la propiedad exhaustiva*

$$A_1 + A_2 \ldots + A_M = \mathsf{S} \tag{A1.7}$$

entonces

$$P(A_1) + P(A_2) + \ldots + P(A_M) = 1 \tag{A1.8}$$

Para probar esta propiedad, utilizamos primero el axioma *i* en la ecuación (A1.7) y de ese modo escribimos

$$P(A_1 + A_2 + \ldots + A_M) = 1$$

Después generalizamos el axioma *iii* escribiendo

$$P(A_1 + A_2 + \ldots + A_M) = P(A_1) + P(A_2) + \ldots + P(A_M)$$

Por tanto, se produce el resultado de la ecuación (A1.8). Cuando los M eventos son *igualmente probables* (es decir, tienen igual probabilidad de ocurrencia), entonces la ecuación (A1.8) se simplifica como

$$P(A_i) = \frac{1}{M}, \quad i = 1, 2, \ldots, M$$

Propiedad 3

Cuando los eventos A y B no son mutuamente excluyentes, entonces la probabilidad del evento unión "A o B" es igual a

$$P(A + B) = P(A) + P(B) - P(AB) \tag{A1.9}$$

donde P(AB) es la probabilidad del evento conjunto "A y B".

La probabilidad $P(AB)$ se denomina *probabilidad conjunta*. Ésta tiene la siguiente interpretación de frecuencia relativa:

$$P(AB) = \lim_{n\to\infty}\left(\frac{N_n(AB)}{n}\right) \tag{A1.10}$$

donde $N_n(AB)$ denota el número de veces que los eventos A y B ocurren simultáneamente en n ensayos del experimento. El axioma *iii* es un caso especial de la ecuación (A1.9); cuando A y B son mutuamente excluyentes, $P(AB)$ es cero y la ecuación (A1.9) se reduce a la misma forma que la ecuación (A1.5).

■ PROBABILIDAD CONDICIONAL

Suponga que realizamos un experimento que implica un par de eventos A y B. Sea $P(B|A)$ la probabilidad del evento B, dado que ha ocurrido el evento A. La probabilidad del evento $P(B|A)$ se denomina la *probabilidad condicional de B dado A*. Suponiendo que A tiene probabilidad distinta de cero, la probabilidad condicional $P(B|A)$ se define mediante

$$P(B|A) = \frac{P(AB)}{P(A)} \tag{A1.11}$$

donde $P(AB)$ es la probabilidad conjunta de A y B.

Justificamos la definición de la probabilidad condicional dada en la ecuación (A1.11) presentando una interpretación de frecuencia relativa de ella. Suponga que efectuamos un experimento y examinamos la ocurrencia de un par de eventos A y B. Dejemos que $N_n(AB)$ denote el número de veces que ocurre el evento conjunto AB en n ensayos. Suponga que en los mismos n ensayos el evento A ocurre $N_n(A)$ veces. Puesto que el evento conjunto AB corresponde tanto a la ocurrencia de A como de B, se concluye que $N_n(A)$ debe incluir $N_n(AB)$. En otras palabras, tenemos

$$\frac{N_n(AB)}{N_n(A)} \leq 1$$

La proporción $N_n(AB)/N_n(A)$ representa la frecuencia relativa de B dado que A ha ocurrido. Para n grande, esta proporción es igual a la probabilidad condicional $P(B|A)$; es decir,

$$P(B|A) = \lim_{n\to\infty}\left(\frac{N_n(AB)}{N_n(A)}\right)$$

o equivalentemente,

$$P(B \mid A) = \lim_{n\to\infty}\left(\frac{N_n(AB)/n}{N_n(A)/n}\right)$$

Admitiendo que

$$P(AB) = \lim_{n\to\infty}\left(\frac{N_n(AB)}{n}\right)$$

y

$$P(A) = \lim_{n\to\infty}\left(\frac{N_n(A)}{n}\right)$$

se obtiene el resultado de la ecuación (A1.11).

Es posible reescribir la ecuación (A1.11) como

$$P(AB) = P(B|A)P(A) \tag{A1.12}$$

Es claro que también podemos escribir

$$P(AB) = P(A|B)P(B) \tag{A1.13}$$

Por tanto, es factible establecer que *la probabilidad conjunta de dos eventos puede expresarse como el producto de la probabilidad condicional de un evento dado el otro, y la probabilidad elemental del otro.* Note que las probabilidades condicionales $P(B \mid A)$ y $P(A \mid B)$ tienen en esencia las mismas propiedades que las diversas probabilidades definidas previamente.

Es posible que existan situaciones en las que la probabilidad condicional $P(A \mid B)$ y las probabilidades $P(A)$ y $P(B)$ se determinen fácilmente en forma directa, aunque se desee la probabilidad condicional $P(B \mid A)$. De las ecuaciones (A1.12) y (A1.13), se concluye que, siempre que $P(A) \neq 0$, podemos determinar $P(B \mid A)$ utilizando la relación

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{A1.14}$$

Esta relación es una forma especial de la *regla de Bayes*.

Suponga que la probabilidad condicional $P(B \mid A)$ es simplemente igual a la probabilidad elemental de ocurrencia del evento B, es decir,

$$P(B|A) = P(B)$$

Bajo esta condición, la probabilidad de ocurrencia del evento conjunto AB es igual al producto de las probabilidades elementales de los eventos A y B:

$$P(AB) = P(A)P(B)$$

de manera que

$$P(A|B) = P(A)$$

Es decir, la probabilidad condicional del evento A, suponiendo la ocurrencia del evento B, es simplemente igual a la probabilidad elemental del evento A. Del mismo modo vemos que en este caso un conocimiento de la ocurrencia de un evento no nos dice más acerca de la probabilidad de la ocurrencia del otro evento de lo que sabíamos sin dicho conocimiento. Los eventos A y B que satisfacen esta condición se dice que son *estadísticamente independientes.*

A1.2 *Variables aleatorias*

Es usual, en particular cuando se usa el lenguaje del espacio muestral, pensar en el resultado de un experimento como una variable que puede moverse por el conjunto de puntos muestrales y cuyo valor está determinado por el experimento. *Una función cuyo dominio es un espacio muestral y cuyo intervalo es algún conjunto de números reales recibe el nombre de variable aleatoria del experimento.* Sin embargo, el término *variable aleatoria* es un poco confuso. En primer lugar, la palabra *aleatoria* no se usa en el sentido de igual probabilidad de ocurrencia, para el cual debe reservarse. En segundo, la palabra *variable,* no implica dependencia (del resultado experimental), lo cual es una parte esencial de significado. El término está tan inmerso en la literatura de la probabilidad que su uso ha persistido.

Cuando el resultado de un experimento es s, la variable aleatoria se denota como $X(s)$ o simplemente X. Por ejemplo, el espacio muestral que representa los resultados del lanzamiento de un dado es un conjunto de seis puntos muestrales que pueden considerarse como los enteros 1, 2, ..., 6. En estas condiciones, si identificamos el punto muestral k con el evento de que k puntos se muestran cuando se lanza el dado, la función $X(k) = k$ es una variable aleatoria tal que $X(k)$ es igual al número de puntos que se exhiben cuando se muestra el dado. En este ejemplo, la variable aleatoria sólo toma un conjunto discreto de valores. En ese caso, decimos que estamos tratando con una *variable aleatoria discreta.* De manera más precisa, la variable aleatoria X puede tomar solamente un número finito de valores en un intervalo de observación finito. Sin embargo, si es posible que la variable X adquiera cualquier valor en un intervalo de observación completa, X recibe el nombre de *variable aleatoria continua.* Por ejemplo, la variable aleatoria que representa la amplitud de un voltaje de ruido en un instante particular es una variable aleatoria continua porque puede tomar cualquier valor entre más y menos infinito.

Para seguir adelante, necesitamos una descripción probabilística de las variables aleatorias que funcione igualmente bien para variables aleatorias discretas y continuas. Vamos a considerar la variable aleatoria X y la probabilidad del evento $X \leq x$. Denotamos esta probabilidad mediante $P(X \leq x)$. Es claro que esta probabilidad es una función de la *variable simulada* x. Para simplificar la notación, escribimos

$$F_X(x) = P(X \leq x) \tag{A1.15}$$

La función $F_X(x)$ se conoce como la *función de distribución acumulativa* o simplemente la *función de distribución* de la variable aleatoria X. Advierta que $F_X(x)$ es una función de x, no de la variable aleatoria X. Sin embargo, depende de la asignación de la variable aleatoria X, lo cual explica el uso de X como subíndice. En cualquier punto x, la función de distribución $F_X(x)$ expresa una probabilidad.

La función de distribución $F_X(x)$ tiene las siguientes propiedades, las cuales siguen directamente de la ecuación (A1.15):

1. La función de distribución $F_X(x)$ está acotada entre cero y uno.
2. La función de distribución $F_X(x)$ es una función no decreciente de x; es decir,

$$F_X(x_1) \leq F_X(x_2) \quad \text{si } x_1 < x_2 \tag{A1.16}$$

Una descripción alternativa de la probabilidad de la variable aleatoria X es a menudo útil. Ésta corresponde a la derivada de la función de distribución, como indica

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{A1.17}$$

la cual se denomina *función de densidad de probabilidad* de la variable aleatoria X. Advierta que la diferenciación en la ecuación (A1.17) es con respecto a la variable simulada x. El nombre, función de densidad, surge del hecho de que la probabilidad del evento $x_1 < X \leq x_2$ es igual a

$$\begin{aligned} P(x_1 < X \leq x_2) &= P(X \leq x_2) - P(X \leq x_1) \\ &= F_X(x_2) - F_X(x_1) \\ &= \int_{x_1}^{x_2} f_X(x)\, dx \end{aligned} \tag{A1.18}$$

La probabilidad de un intervalo es entonces el área bajo la función de densidad de probabilidad en ese intervalo. Dejando $x_1 = -\infty$ en la ecuación (A1.18), y cambiando un poco la notación, vemos de inmediato que la función de distribución se define en términos de la función de densidad de probabilidad del modo siguiente:

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi \tag{A1.19}$$

Puesto que $F_X(\infty) = 1$, correspondiendo a la probabilidad de un cierto evento, y $F_X(-\infty) = 0$, relativa a la probabilidad de un evento imposible, determinamos fácilmente a partir de la ecuación (A1.18) que

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \tag{A1.20}$$

Mencionamos antes que una función de distribución siempre debe ser no decreciente, lo cual significa que su derivada o la función de densidad de probabilidad siempre deben ser no negativas. Por tanto, es posible establecer que *una función de densidad de probabilidad siempre debe ser una función no negativa y con un área total de uno.*

Hasta ahora hemos centrado la atención en situaciones que implican una sola variable aleatoria. Sin embargo, encontramos a menudo que el resultado de un experimento requiere para su descripción diversas variables aleatorias. Consideramos ahora situaciones que incluyen dos variables aleatorias. La descripción probabilística desarrollada de esta manera puede extenderse sin ninguna dificultad a cualquier número de variables aleatorias.

Considere dos variables aleatorias X y Y. Definimos *la función de distribución conjunta* $F_{X,Y}(x, y)$ *como la probabilidad de que la variable aleatoria* X *sea menor o igual que un valor especificado* x *y que la variable aleatoria* Y *sea menor o igual que un valor especificado* y. Es posible que las variables X y Y sean dos variables aleatorias unidimensionales independientes o las componentes de una sola variable aleatoria bidimensional. En cualquier caso, el espacio muestral conjunto es el plano xy. La función de distribución conjunta $F_{X,Y}(x, y)$ es la probabilidad de que el resultado de un experimento originará un punto muestral que se encuentra dentro del cuadrante $(-\infty < X = x, -\infty < Y \leq y)$ del espacio muestral conjunto. Es decir,

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) \tag{A1.21}$$

Suponga que la función de distribución conjunta $F_{X,Y}(x, y)$ es continua en todos lados, y que la derivada parcial

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \tag{A1.22}$$

existe y es continua en todas partes. Denominamos a la función $f_{X,Y}(x, y)$ la *función de densidad de probabilidad conjunta* de las variables aleatorias X y Y. La función de distribución conjunta

$F_{X,Y}(x, y)$ es una función no decreciente tanto de x como de y. Por lo tanto, a partir de la ecuación (A1.22) se concluye que la función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$ siempre es no negativa. Asimismo, el volumen total bajo la gráfica de una función de densidad de probabilidad conjunta debe ser la unidad, según indica

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(\xi,\eta)\,d\xi\,d\eta = 1 \tag{A1.23}$$

La función de densidad de probabilidad para una sola variable aleatoria (X por ejemplo) puede obtenerse de su función de densidad de probabilidad conjunta con una segunda variable aleatoria (Y por ejemplo). Advierta primero que

$$F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(\xi,\eta)\,d\xi\,d\eta \tag{A1.24}$$

En consecuencia, al diferenciar ambos miembros de la ecuación (A1.24) con respecto a x, obtenemos la relación deseada:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,\eta)\,d\eta \tag{A1.25}$$

De ese modo la función de densidad de probabilidad $f_X(x)$ se obtiene de la función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$ integrándola simplemente sobre todos los valores posibles de la variable aleatoria no deseada Y. El uso de argumentos similares en la otra dimensión produce $f_Y(y)$. Las funciones de densidad de probabilidad $f_X(x)$ y $f_Y(y)$ reciben el nombre de *densidades marginales*. Por consiguiente, la función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$ contiene toda la información posible acerca de las variables aleatorias X y Y.

Suponga que X y Y son dos variables aleatorias continuas con función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$. La *función de densidad de probabilidad condicional* de Y dado que $X = x$ está definida por

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{A1.26}$$

siempre que $f_X(x) > 0$, donde $f_X(x)$ es la densidad marginal de X. Es posible considerar la función $f_Y(y|x)$ como una función de la variable y, con la variable x arbitraria, aunque *fija*, por lo que ésta satisface todos los requerimientos de una función de probabilidad ordinaria, como se expresa mediante

$$f_Y(y|x) \geq 0$$

y

$$\int_{-\infty}^{\infty} f_Y(y|x)\,dy = 1$$

Si las variables aleatorias X y Y *son estadísticamente independientes*, entonces el conocimiento del resultado de X no puede de ninguna forma afectar la distribución de Y. El resultado es que la función de densidad de probabilidad condicional $f_Y(y|x)$ se reduce a la densidad marginal $f_Y(y)$, como indica

$$f_Y(y|x) = f_Y(y)$$

En ese caso, podemos expresar la función de densidad de probabilidad conjunta de las variables aleatorias X y Y como el producto de sus respectivas densidades marginales, de acuerdo con

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

En palabras, es posible establecer que si la función de densidad de probabilidad conjunta de las variables aleatorias X y Y es igual al producto de sus densidades marginales, entonces X y Y son estadísticamente independientes.

A1.3 *Promedios estadísticos*

Luego de haber explicado la probabilidad y algunas de sus ramificaciones, buscamos ahora formas para determinar el *comportamiento promedio* y los resultados que surgen en los experimentos aleatorios.

El *valor esperado* o *media* de una variable aleatoria X está definido por

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{A1.27}$$

donde E denota el operador de *esperanza estadística*. Es decir, la media μ_X ubica el centro de gravedad del área bajo la curva de densidad de probabilidad de la variable aleatoria X. Para interpretar el valor esperado μ_X, escribimos la integral en la ecuación de definición (A1.27) como el límite de una suma aproximativa formulado del modo siguiente. Sea $\{x_k | k = 0, \pm 1, \pm 2, \ldots\}$ un conjunto de puntos espaciados uniformemente sobre la línea real:

$$x_k = \left(k + \frac{1}{2}\right)\Delta, \quad k = 0, \pm 1, \pm 2, \ldots \tag{A1.28}$$

donde Δ es el espaciamiento entre puntos adyacentes. Podemos entonces reescribir la ecuación (A1.27) como la forma límite de una suma:

$$\begin{aligned} E[X] &= \lim_{\Delta \to 0} \sum_{k=-\infty}^{\infty} \int_{k\Delta}^{(k+1)\Delta} x_k f_X(x)\, dx \\ &= \lim_{\Delta \to 0} \sum_{k=-\infty}^{\infty} x_k P\left(x_k - \frac{\Delta}{2} < X \le x_k + \frac{\Delta}{2}\right) \end{aligned} \tag{A1.29}$$

Para una interpretación física de la suma del segundo miembro de la ecuación (A1.29), suponga que efectuamos n observaciones independientes de la variable aleatoria X. Considere que $N_n(k)$ denota el número de veces que la variable aleatoria X cae dentro del k-ésimo cajón:

$$x_k - \frac{\Delta}{2} < X \le x_k + \frac{\Delta}{2}, \quad k = 0, \pm 1, \pm 2, \ldots$$

En ese caso, cuando el número de observaciones, n, se hace grande, la proporción $N_n(k)/n$ tiende a la probabilidad $P(x_k - \Delta/2 < X \le x_k + \Delta/2)$. En consecuencia, es posible aproximar el valor esperado de la variable aleatoria X como

$$\begin{aligned} E[X] &\simeq \sum_{k=-\infty}^{\infty} x_k \left(\frac{N_n(k)}{n}\right) \\ &= \frac{1}{n} \sum_{k=-\infty}^{\infty} x_k N_n(k), \qquad n \text{ grande} \end{aligned} \tag{A1.30}$$

Admitimos ahora a la cantidad en el segundo miembro de la ecuación (A1.30) simplemente como el *promedio muestral*. La suma se toma sobre todos los valores de x_k, cada uno de los cuales se pondera por medio del número de veces que ocurre; la suma se divide entonces entre el número total de observaciones para producir el promedio muestral. En realidad, la ecuación (A1.30) proporciona las bases para el cálculo del valor esperado $E[X]$.

A continuación consideramos una situación más general. Sea X una variable aleatoria y $g(X)$ una función de X definida sobre la línea real. La cantidad que se obtiene al dejar que el argumento de la función $g(X)$ sea una variable aleatoria es también una variable aleatoria, lo cual denotamos como

$$Y = g(X) \tag{A1.31}$$

Para encontrar el valor esperado de la variable aleatoria Y, podríamos encontrar desde luego la función de densidad de probabilidad $f_Y(y)$ y aplicar después la fórmula estándar

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Sin embargo, un procedimiento más simple es escribir

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \tag{A1.32}$$

En realidad, la ecuación (A1.32) puede considerarse como la generalización del concepto del valor esperado para una función arbitraria $g(X)$ de una variable aleatoria X.

■ MOMENTOS

Para el caso especial $g(X) = X^n$, utilizando la ecuación (A1.32) obtenemos el *n*-ésimo *momento* de la función de distribución de probabilidad de la variable aleatoria X; es decir,

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\, dx \tag{A1.33}$$

Por mucho los momentos más importantes de X son los primeros dos. De modo que al poner $n = 1$ en la ecuación (A1.33) se genera la media de la variable aleatoria, como se muestra en la ecuación (A1.27), en tanto que al dejar $n = 2$ se produce el *valor cuadrático medio de* X.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \tag{A1.34}$$

Es posible también definir los *momentos centrales* que son simplemente los momentos de la diferencia entre una variable aleatoria X y su media μ_X. Por tanto, el momento central enésimo es

$$E[(X - \mu_X)^n] = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x)\, dx \tag{A1.35}$$

Desde luego, para $n = 1$ el momento central corresponde a cero, en tanto que para $n = 2$ el segundo momento central se conoce como la *varianza* de la variable aleatoria X, la cual se escribe como

$$\text{var}[X] = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx \tag{A1.36}$$

La varianza de una variable aleatoria X se denota comúnmente como σ_X^2. La raíz cuadrada de la varianza, a saber, σ_X, recibe el nombre de *desviación estándar de la variable aleatoria* X.

La varianza σ_X^2 de una variable aleatoria X es en algún sentido una medida de la "aleatoriedad" de la variable. Al especificar la varianza σ_X^2, restringimos esencialmente el ancho efectivo de la función de densidad de probabilidad $f_X(x)$ de la variable aleatoria X en torno a la

media μ_X. Un enunciado preciso de esta restricción se debe a Chebyshev. La *desigualdad de Chebyshev* establece que para cualquier número positivo ε, tenemos

$$P(|X-\mu_X| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2} \tag{A1.37}$$

A partir de esta desigualdad vemos que la media y la varianza de una variable aleatoria ofrecen una *descripción parcial* de su distribución de probabilidad, de ahí su uso común en la práctica.

Observamos a partir de las ecuaciones (A1.34) y (A1.36) que la varianza σ_X^2 y el valor cuadrático medio $E[X^2]$ se relacionan por medio de

$$\begin{aligned}\sigma_X^2 &= E[X^2 - 2\mu_X X + \mu_X^2] \\ &= E[X^2] - 2\mu_X E[X] + \mu_X^2 \\ &= E[X^2] - \mu_X^2\end{aligned} \tag{A1.38}$$

donde, en la segunda línea hemos utilizado la propiedad de *linealidad* del operador de esperanza estadística. La ecuación (A1.38) demuestra que si la media μ_X es cero, entonces la varianza σ_X^2 y el valor $E[X^2]$ de la variable aleatoria X son iguales.

Función característica

Otro promedio estadístico importante es la *función característica* $\phi_X(v)$ de la distribución de probabilidad de la variable aleatoria X, la cual se define como la esperanza de la función exponencial compleja $\exp(jvX)$, como se indica mediante

$$\begin{aligned}\phi_X(v) &= E[\exp(jvX)] \\ &= \int_{-\infty}^{\infty} f_X(x)\exp(jvx)\,dx\end{aligned} \tag{A1.39}$$

donde v es real y $j = \sqrt{-1}$. En otras palabras, la función característica $\phi_X(v)$ es (excepto por un cambio de signo en el exponente) la transformada de Fourier de la función de densidad de probabilidad $f_X(x)$; la transformada de Fourier se repasa en el apéndice 2. En esta relación hemos utilizado $\exp(jvx)$ en lugar de $\exp(-jvx)$, de modo que haya concordancia con la convención adoptada en la teoría de la probabilidad. Admitiendo que v y x desempeñan papeles análogos a los de las variables $2\pi f$ y t de las transformadas de Fourier, respectivamente, deducimos la siguiente relación inversa a partir de la analogía con la transformada inversa de Fourier:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(v)\exp(-jvx)\,dv \tag{A1.40}$$

Esta relación puede utilizarse para evaluar la función de densidad de probabilidad $f_X(x)$ de la variable aleatoria X a partir de su función característica $\phi_X(v)$.

Momentos conjuntos

Considere a continuación un par de variables aleatorias X y Y. Un conjunto de promedios estadísticos de importancia en este caso es el de los *momentos conjuntos*, a saber, el valor esperado de X^iY^k, donde i y k pueden asumir cualesquiera valores enteros positivos. De ese modo es posible escribir

$$E[X^iY^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i y^k f_{X,Y}(x,y)\,dx\,dy \tag{A1.41}$$

Un momento conjunto de importancia particular es la *correlación* definida por $E[XY]$, que corresponde a $i = k = 1$ en la ecuación (A1.41).

La correlación de las variables aleatorias centradas $X - E[X]$ y $Y - E[Y]$, es decir, el momento conjunto

$$\text{cov}[XY] = E[(X - E[X])(Y - E[Y])] \tag{A1.42}$$

recibe el nombre de la *covarianza* de X y Y. Dejando $\mu_X = E[X]$ y $\mu_y = E[Y]$, podemos expandir la ecuación (A1.42) para obtener el resultado

$$\text{cov}[XY] = E[XY] - \mu_X \mu_Y \tag{A1.43}$$

Considere que σ_X^2 y σ_Y^2 denotan las varianzas de X y Y, respectivamente; entonces la covarianza de X y Y, normalizada con respecto a $\sigma_X \sigma_Y$, se denomina el *coeficiente de correlación* de X y Y:

$$\rho = \frac{\text{cov}[XY]}{\sigma_X \sigma_Y} \tag{A1.44}$$

Afirmamos que las dos variables aleatorias X y Y están *no correlacionadas* si y sólo si su covarianza es cero, es decir, si y sólo si

$$\text{cov}[XY] = 0$$

Decimos que son *ortogonales* si y sólo si su correlación es cero, es decir, si y sólo si

$$E[XY] = 0$$

A partir de la ecuación (A1.43) observamos que si una de las variables aleatorias X y Y o ambas tienen medias cero, y si ellas son ortogonales, entonces están no correlacionadas, y viceversa. Advierta también que si X y Y son estadísticamente independientes, entonces son no correlacionadas; sin embargo, lo inverso de este enunciado no necesariamente es cierto.

APÉNDICE 2

REPRESENTACIÓN DE SEÑALES Y SISTEMAS

A2.1 *Análisis de Fourier*

Sea $g(t)$ una *señal determinística no periódica*, expresada como alguna función del tiempo t. Por definición, la *transformada de Fourier* de la señal $g(t)$ está dada por la integral

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp(-j2\pi ft)\, dt \tag{A2.1}$$

donde $j = \sqrt{-1}$, y la variable f denota la *frecuencia*. Dada la transformada de Fourier $G(f)$, la señal original $g(t)$ se recupera exactamente utilizando la fórmula de la *transformada inversa de Fourier*:

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft)\, df \tag{A2.2}$$

Advierta que en las ecuaciones (A2.1) y (A2.2) hemos usado una letra minúscula para denotar la función en el tiempo y una letra mayúscula para indicar la correspondiente función en frecuencia. Las funciones $g(t)$ y $G(f)$ se dice que constituyen un par de transformadas de Fourier.

Para que exista una transformada de Fourier $g(t)$, es suficiente aunque no necesario que $g(t)$ cumpla tres condiciones conocidas colectivamente como *condiciones de Dirichlet*:

1. La función $g(t)$ es univaluada, con un número finito de máximos y mínimos en cualquier intervalo de tiempo finito.
2. La función $g(t)$ tiene un número finito de discontinuidades en todo intervalo de tiempo finito.
3. La función $g(t)$ es absolutamente integrable, es decir,

$$\int_{-\infty}^{\infty} |g(t)|\, dt < \infty$$

Podemos ignorar sin ningún problema la cuestión de la existencia de la transformada de Fourier de una función del tiempo $g(t)$ cuando ésta es una descripción especificada con exactitud de una señal físicamente realizable. En otras palabras, la realizabilidad física es una condición suficiente para la transformada de Fourier. En realidad, podemos ir un paso adelante y establecer que todas las señales de energía, es decir, señales $g(t)$ para las cuales

$$\int_{-\infty}^{\infty} |g(t)|^2 dt < \infty$$

son transformables de Fourier.

La transformada de Fourier proporciona la herramienta matemática para medir el contenido de frecuencia o espectro de una señal. Por esta razón, los términos *transformada de Fourier* y *espectro* se emplean a menudo de manera indistinta. De ese modo, dada una señal $g(t)$ con transformada de Fourier $G(f)$, es posible referirnos a $G(f)$ como el espectro de la señal $g(t)$. Por las mismas razones nos referimos a $|G(f)|$ como el *espectro de magnitud* de la señal $g(t)$, y a $\arg\{G(f)\}$ como su *espectro de fase*.

■ PROPIEDADES DE LA TRANSFORMADA DE FOURIER

Es útil tener información acerca de la relación entre la función en el tiempo $g(t)$ y su transformada de Fourier $G(f)$, así como en relación con los efectos que las diversas operaciones sobre la función $g(t)$ tienen sobre la transformada $G(f)$. Lo anterior puede lograrse examinando ciertas propiedades de la transformada de Fourier, las cuales se resumen en la tabla A6.2.

■ FUNCIÓN DELTA DE DIRAC

En términos generales, la teoría de la transformada de Fourier es aplicable sólo a funciones en el tiempo que satisfacen las condiciones de Dirichlet. Tales funciones incluyen señales de energía. Sin embargo, sería bastante deseable extender esta teoría por dos vías:

1. Combinar la serie de Fourier y la transformada de Fourier en una teoría unificada, de manera que la serie de Fourier pueda tratarse como un caso especial en la transformada de Fourier.
2. Incluir señales de potencia (es decir, señales para las cuales la potencia promedio es finita) en la lista de señales para las cuales podemos aplicar la transformada de Fourier.

Resulta que estos objetivos pueden cumplirse mediante el "uso apropiado" de la *función delta de Dirac*, o *impulso unitario*.

La función delta de Dirac o función delta, denotada por $\delta(t)$, se define como aquella que tiene amplitud cero en todas partes salvo en $t = 0$, donde es infinitamente grande, de tal modo que contiene un área unitaria bajo su curva, es decir,

$$\delta(t) = 0, \quad t \neq 0 \tag{A2.3}$$

y

$$\int_{-\infty}^{\infty} \delta(t)\, dt = 1 \tag{A2.4}$$

Una implicación de este par de relaciones es que la función $\delta(t)$ debe ser una función par del tiempo t que está centrada en $t = 0$.

Para que la función delta tenga significado, tiene que aparecer como un factor en el integrando de una integral con respecto al tiempo y, en ese caso, en términos generales, únicamente cuando el otro factor en el integrando sea una función continua del tiempo. Sea $g(t)$ una función de tales características y considere el producto de $g(t)$ y la función delta recorrida en el tiempo $\delta(t - t_0)$. De acuerdo con las dos ecuaciones de definición, ecuaciones (A2.3) y (A2.4), es posible expresar la integral de este producto del modo siguiente:

$$\int_{-\infty}^{\infty} g(t)\, \delta(t - t_0)\, dt = g(t_0) \tag{A2.5}$$

La operación que se indica en el primer miembro de esta ecuación selecciona el valor $g(t_0)$ de la función $g(t)$ en el tiempo $t = t_0$, donde $-\infty < t < \infty$. Por tanto, La ecuación (A2.5) se conoce como la *propiedad de selección* de la función delta. Esta propiedad se utiliza algunas veces como la ecuación de definición de una función delta; en efecto, incorpora las ecuaciones (A2.3) y (A2.4) en una sola relación.

Advirtiendo que la función delta $\delta(t)$ es una función par de t, podemos reescribir la ecuación (A2.5) de manera que se subraye su semejanza con la integral de convolución, como se indica mediante

$$\int_{-\infty}^{\infty} g(\tau)\, \delta(t - \tau)\, d\tau = g(t) \tag{A2.6}$$

En palabras, la convolución de cualquier función con la función delta deja sin cambio esa función. Nos referiremos a este enunciado como la *propiedad de replicación* de la función delta.

Es importante darse cuenta de que ninguna función en el sentido ordinario tiene las propiedades de las ecuaciones (A2.3) y (A2.4) o la propiedad de selección equivalente de la ecuación (A2.5). Sin embargo, es posible imaginar una secuencia de funciones que tiene picos progresivamente más altos y más delgados en $t = 0$, con el área bajo la curva permaneciendo igual que la unidad, en tanto que el valor de la función tiende a cero en cualquier punto salvo en $t = 0$, donde tiende a infinito. Es decir, podemos considerar la función delta como *la forma límite de un pulso de área unitaria cuando la duración del pulso tiende a cero.* No tiene importancia qué tipo de forma de pulso se utiliza.

■ Transformadas de Fourier de señales periódicas

Es bien sabido que al utilizar la serie de Fourier, una señal periódica puede representarse como una suma de exponenciales complejas. Asimismo, en un sentido limitado, es posible definir las transformadas de Fourier mediante exponenciales complejas. Por lo tanto, parece razonable representar una señal periódica en términos de una transformada de Fourier, siempre y cuando se permita que esta transformación incluya funciones delta.

Considere a continuación una señal periódica $g_{T_0}(t)$ *de periodo* T_0. Podemos representar $g_{T_0}(t)$ en términos de la *serie de Fourier exponencial compleja*:

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_0 t) \tag{A2.7}$$

donde c_n es el *coeficiente de Fourier complejo* definido por

$$c_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_{T_0}(t)\exp(-j2\pi n f_0)\,dt \tag{A2.8}$$

y f_0 es la *frecuencia fundamental* definida como el recíproco del periodo T_0; es decir,

$$f_0 = \frac{1}{T_0} \tag{A2.9}$$

Sea $g(t)$ una función similar a un pulso, que es igual a $g_{T_0}(t)$ sobre un periodo y cero en cualquier otro lado; es decir,

$$g(t) = \begin{cases} g_{T_0}(t), & -\dfrac{T_0}{2} < t \le \dfrac{T_0}{2} \\ 0 & \text{en otro lado} \end{cases} \tag{A2.10}$$

La señal periódica $g_{T_0}(t)$ puede expresarse ahora en términos de la función $g(t)$ como una sumatoria infinita, según indica

$$g_{T_0}(t) = \sum_{m=-\infty}^{\infty} g(t - mT_0) \tag{A2.11}$$

Con base en esta representación, podemos ver $g(t)$ como una *función generatriz*, que genera la señal periódica $g_{T_0}(t)$.

La función $g(t)$ es transformable de Fourier. Por tanto, es factible reescribir la fórmula para el coeficiente de Fourier complejo del siguiente modo:

$$\begin{aligned} c_n &= f_0 \int_{-\infty}^{\infty} g(t)\exp(-j2\pi n f_0 t)\,dt \\ &= f_0 G(n f_0) \end{aligned} \tag{A2.12}$$

donde $G(nf_0)$ es la transformada de Fourier de $g(t)$ evaluada a la frecuencia nf_0. Consecuentemente podemos reescribir la fórmula para la reconstrucción de la señal periódica $g_{T_0}(t)$ como

$$g_{T_0}(t) = f_0 \sum_{n=-\infty}^{\infty} G(nf_0) \exp(j2\pi nf_0 t) \tag{A2.13}$$

o, de modo equivalente, de acuerdo con la ecuación (A2.11)

$$\sum_{m=-\infty}^{\infty} g(t - mT_0) = f_0 \sum_{n=-\infty}^{\infty} G(nf_0) \exp(j2\pi nf_0 t) \tag{A2.14}$$

La ecuación (A2.14) es una forma de la *fórmula de Poisson.*

Es interesante observar que la función $g(t)$, la cual constituye un periodo de la señal periódica $g_{T_0}(t)$, tiene un espectro continuo definido por la señal $G(f)$. Por otro lado, la propia señal periódica $g_{T_0}(t)$ tiene un espectro discreto. Por tanto, concluimos que la *periodicidad en el dominio del tiempo tiene el efecto de cambiar la descripción o espectro en el dominio de la frecuencia de la señal en una forma discreta definida en múltiplos enteros de la frecuencia fundamental.*

■ PARES DE TRANSFORMADAS DE FOURIER

La tabla A6.3 presenta una lista de algunos pares de transformadas de Fourier utilizados comúnmente, cuya derivación sigue del material que acaba de presentarse.

■ TRANSMISIÓN DE SEÑALES A TRAVÉS DE SISTEMAS LINEALES

Un *sistema* se refiere a cualquier dispositivo físico que produce una señal de salida en respuesta a una señal de entrada. Es usual referirse a la señal de entrada como la *excitación* y a la señal de salida como la *respuesta.* En un sistema *lineal,* se cumple el *principio de superposición*; es decir, la respuesta de un sistema lineal a varias excitaciones aplicadas en forma simultánea es igual a la suma de las respuestas del sistema cuando cada excitación se aplica individualmente.

En el dominio del tiempo, un sistema lineal se describe en términos de su *respuesta al impulso,* la cual se define como *la respuesta del sistema (con condiciones iniciales cero) a un impulso unitario o función delta $\delta(t)$ aplicado a la entrada del sistema.* Si el sistema es *invariante con el tiempo,* entonces la forma de la respuesta al impulso es la misma sin que importe cuándo se aplicó al sistema el impulso unitario. Por tanto, suponiendo que se aplica el impulso unitario o la función delta en el tiempo $t = 0$, podemos denotar la respuesta al pulso de un sistema lineal invariante con el tiempo por medio de $h(t)$. Considere que el sistema está sujeto a una excitación arbitraria $x(t)$. La respuesta $y(t)$ del sistema se define en términos de la respuesta al impulso $h(t)$ por medio de

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)\, d\tau \tag{A2.15}$$

la cual recibe el nombre de *integral de convolución.* Equivalentemente, es posible escribir

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)\, dt \tag{A2.16}$$

Por tanto, la convolución es *conmutativa.*

En la integral de convolución están implicadas tres escalas diferentes: *tiempo de excitación* τ, *tiempo de respuesta* t y *tiempo de memoria del sistema* $t - \tau$. Esta relación constituye la base del análisis en el dominio del tiempo de sistemas lineales invariantes con el tiempo. De acuerdo con la ecuación (A2.15), el valor presente de la respuesta de un sistema lineal invariante con el tiempo es una integral ponderada sobre la historia pasada de la señal de entrada, ponderada a su

vez de acuerdo con la respuesta al impulso del sistema. De esa manera la respuesta al impulso actúa como una *función de memoria* del sistema.

Respuesta en frecuencia de sistemas lineales invariantes con el tiempo

Considere un sistema lineal invariante con el tiempo de respuesta $h(t)$ excitado por una entrada exponencial compleja de amplitud unitaria y frecuencia f, es decir,

$$x(t) = \exp(j2\pi ft)$$

Utilizando esta excitación en la ecuación (A2.16), la respuesta del sistema se obtiene como

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) \exp[j2\pi f(t-\tau)]\, d\tau \\ &= \exp(j2\pi ft) \int_{-\infty}^{\infty} h(\tau) \exp(-j2\pi f\tau)\, d\tau \end{aligned} \tag{A2.17}$$

Defina la *respuesta en frecuencia* del sistema, como la transformada de Fourier de su respuesta al impulso, según indica

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp(-j2\pi ft)\, dt \tag{A2.18}$$

La integral en la última línea de la ecuación (A2.17) es la misma que la de la ecuación (A2.18), salvo que se usa τ en lugar de t. Por lo tanto, podemos escribir la ecuación (A2.17) en la forma

$$y(t) = H(f) \exp(j2\pi ft) \tag{A2.19}$$

La respuesta de un sistema lineal invariante con el tiempo a una función exponencial compleja de frecuencia f es, por tanto, la misma función exponencial compleja multiplicada por un coeficiente constante $H(f)$.

La respuesta en frecuencia $H(f)$ es, en general, una cantidad compleja, de modo que es viable expresarla

$$H(f) = |H(f)| \exp[j\beta(f)] \tag{A2.20}$$

donde $|H(f)|$ recibe el nombre de la *respuesta en magnitud*, y $\beta(f)$ es la *fase* o *respuesta en fase*. En el caso especial de un sistema lineal con una respuesta al impulso de valores reales $h(t)$, la respuesta en frecuencia $H(f)$ exhibe simetría conjugada, lo que quiere decir que

$$|H(f)| = |H(-f)|$$

y

$$\beta(f) = -\beta(-f)$$

Es decir, la respuesta en magnitud $|H(f)|$ de un sistema lineal con respuesta al impulso de valores reales es una función par de la frecuencia, en tanto que la fase $\beta(f)$ es una función impar de la frecuencia.

En algunas aplicaciones resulta preferible trabajar con el logaritmo de $H(f)$ expresado en forma polar, en vez de hacerlo con la propia $H(f)$. Defina el logaritmo natural

$$\log H(f) = \alpha(f) + j\beta(f) \tag{A2.21}$$

donde

$$\alpha(f) = \log|H(f)| \tag{A2.22}$$

La función $\alpha(f)$ se conoce como la *ganancia* del sistema. Ésta se mide en *nepers*, en tanto que $\beta(f)$ se mide en *radianes*. La ecuación (A2.21) indica que la ganancia $\alpha(f)$ y la fase $\beta(f)$ son las partes real e imaginaria del logaritmo (natural) de la respuesta en frecuencia $H(f)$, respectivamente. La ganancia también puede expresarse en decibeles (dB) utilizando la definición

$$\alpha'(f) = 20 \log_{10}|H(f)| \tag{A2.23}$$

Las dos funciones de ganancia $\alpha(f)$ y $\alpha'(f)$ se relacionan por medio de

$$\alpha'(f) = 8.69\alpha(f) \tag{A2.24}$$

Es decir, 1 neper es igual a 8.69 dB.

A2.2 *Ancho de banda*

Las descripciones en el dominio del tiempo y en el dominio de la frecuencia de una señal se relacionan *inversamente*. En particular, podemos indicar los siguientes enunciados importantes:

1. Si cambia la descripción en el dominio del tiempo de una señal, la descripción de la señal en el dominio de la frecuencia cambia de manera *inversa*, y viceversa. Esta relación inversa evita especificaciones arbitrarias de una señal en ambos dominios. En otras palabras, *podemos especificar una función arbitraria del tiempo o un espectro arbitrario, pero no podemos especificarlos a los dos juntos.*
2. Si una señal está estrictamente limitada en frecuencia, la descripción de la señal en el dominio del tiempo seguirá indefinidamente, aun cuando sea posible que su amplitud asuma un valor progresivamente más pequeño. Afirmamos que una señal está *estrictamente limitada en frecuencia* o *estrictamente limitada en banda* si su transformada de Fourier es exactamente cero fuera de una banda finita de frecuencia. El pulso senc

$$\text{senc}(t) = \frac{\text{sen}(\pi t)}{\pi t}$$

es un ejemplo de una señal estrictamente limitada en banda. También está *limitada asintóticamente en el tiempo*, lo cual confirma el enunciado inicial que hicimos para una señal estrictamente limitada en banda. De manera inversa, si la señal está *estrictamente limitada en el tiempo* (es decir, la señal es exactamente cero fuera de un intervalo finito), entonces el espectro de la señal es infinito en extensión, incluso cuando el espectro de amplitud pueda asumir un valor progresivamente más pequeño. Este comportamiento se ejemplifica por medio de un pulso rectangular. Por tanto, podemos establecer que una *señal no puede estar estrictamente limitada tanto en tiempo como en frecuencia.*

El *ancho de banda* de una señal proporciona una medida del *grado de contenido espectral significativo de la señal para frecuencias positivas*. Cuando la señal está estrictamente limitada en banda, el ancho de banda está bien definido. Por ejemplo, el pulso senc $(2Wt)$ tiene un ancho de banda igual a W. Sin embargo, cuando la señal no está limitada estrictamente en banda, como es generalmente el caso, encontramos dificultades al definir el ancho de banda de la señal. El problema surge debido a que el significado de "significativo" asociado al contenido espectral de la señal es matemáticamente impreciso. En consecuencia, no hay una definición aceptada en forma universal del ancho de banda. A pesar de eso, hay definiciones que se usan en forma común para el ancho de banda, como se explica a continuación.

Cuando el espectro de una señal es simétrico con un *lóbulo principal* acotado por *nulos* (es decir, frecuencias a las cuales el espectro es cero) perfectamente definidos, podemos utilizar dicho lóbulo como la base para definir el ancho de banda de la señal. De modo específico, si la señal es *pasobajas* (es decir, su contenido espectral se centra en torno al origen), el ancho de banda se define como la mitad del ancho total del lóbulo espectral principal, puesto que sólo una mitad de

este lóbulo se encuentra dentro de la región de frecuencia positiva. Por ejemplo, un pulso rectangular de duración T segundos tiene un lóbulo espectral principal de ancho total $2/T$ hertz centrado en el origen. En consecuencia, es posible definir el ancho de banda de este pulso rectangular como $1/T$ hertz. Si, por otra parte, la señal es *pasobanda* con lóbulos espectrales principales centrados alrededor de $\pm f_c$, donde f_c es bastante grande, el ancho de banda se define como el ancho del lóbulo principal para frecuencias positivas. Esta definición del ancho de banda recibe el nombre de *ancho de banda de nulo a nulo.* Por ejemplo, un pulso de RF de duración T segundos y frecuencia f_c tiene lóbulos espectrales de ancho $2/T$ hertz centrados alrededor de $\pm f_c$, donde se supone que f_c es grande comparada con $1/T$. Por consiguiente, podemos definir el ancho de banda de nulo a nulo de este pulso de RF como $2/T$ hertz. Con base en las definiciones presentadas aquí, es factible establecer que el corrimiento del contenido espectral de una señal pasobajas mediante una frecuencia suficientemente grande tiene el efecto de duplicar el ancho de banda de la señal; una traslación de frecuencia de este tipo se consigue utilizando modulación.

Otra definición popular del ancho de banda corresponde al *ancho de banda de 3 dB.* Específicamente, si la señal es pasobajas, el ancho de banda de 3 dB se define como la separación entre la frecuencia cero, donde el espectro de magnitud alcanza su valor pico, y la *frecuencia positiva,* a la cual el espectro de amplitud disminuye a $1/\sqrt{2}$ de su valor pico. Por ejemplo, la exponencial que decae $\exp(-at)$ tiene un ancho de banda de 3 dB de $a/2\pi$ hertz. Si, por otro lado, la señal es pasobanda, centrada en $\pm f_c$, el ancho de banda de 3 dB se define como la separación (a lo largo del eje de frecuencias positivas) entre las dos frecuencias a las cuales el espectro de magnitud de la señal cae hasta $1/\sqrt{2}$ del valor pico de f_c. El ancho de banda de 3 dB tiene la ventaja de que puede leerse directamente en una gráfica del espectro de magnitud. Sin embargo, presenta la desventaja de que puede ser engañoso si el espectro de magnitud tiene colas que disminuyen lentamente.

Incluso otra medida para el ancho de banda de una señal es el *ancho de banda del valor cuadrático medio (rms),* que se define como la raíz cuadrada del segundo momento de una forma normalizada propiamente del espectro de magnitud al cuadrado de la señal en torno a un punto elegido en forma adecuada. Suponemos que la señal es pasobajas, por lo que el segundo momento puede tomarse alrededor del origen. Con respecto a la forma normalizada del espectro de magnitud al cuadrado, utilizamos la función no negativa

$$\frac{|G(f)|^2}{\int_{-\infty}^{\infty}|G(f)|^2\,df}$$

en la cual el denominador aplica la normalización correcta en el sentido de que el valor del integrando de este cociente sobre el eje de frecuencia completo es la unidad. De ese modo es posible definir formalmente el ancho de banda rms de una señal pasobajas $g(t)$ con transformada de Fourier $G(f)$ del modo siguiente:

$$W_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} f^2|G(f)|^2\,df}{\int_{-\infty}^{\infty}|G(f)|^2\,df}\right)^{1/2} \tag{A2.25}$$

Una característica atractiva del ancho de banda rms W_{rms} es que éste se presta por sí mismo de una manera más fácil para la evaluación matemática de las otras dos definiciones del ancho de banda, aunque no se mide fácilmente en el laboratorio.

■ Producto tiempo-ancho de banda

Para cualquier familia de señales de pulso que difieren por un factor de escalamiento en el tiempo, el producto de la duración de la señal y su ancho de banda siempre es una constante:

$$(\text{duración} \times \text{ancho de banda}) = \text{constante}$$

Éste recibe el nombre de *producto tiempo-ancho de banda* o *producto ancho de banda-duración*. La constancia del producto tiempo-ancho de banda es otra manifestación de la relación inversa que existe entre las descripciones en el dominio del tiempo y en el dominio de la frecuencia de una señal. En particular, si la duración de una señal pulso se reduce al disminuir la escala de tiempo por un factor a, la escala de frecuencia del espectro de la señal y, por tanto, el ancho de banda de la misma, aumenta en el mismo factor, en virtud de la *propiedad de escalamiento del tiempo* de la transformada de Fourier, por lo que el producto tiempo-ancho de banda de la señal se mantiene constante; vea la entrada 2 de la tabla A6.2. Por ejemplo, un pulso rectangular de duración T segundos tiene un ancho de banda (definido con base en la parte de la frecuencia positiva del lóbulo principal) igual a $1/T$ hertz, lo que hace que el producto tiempo-ancho de banda sea igual a la unidad. Cualquiera que sea la definición que utilicemos para el ancho de banda de una señal, el producto tiempo-ancho de banda permanece constante sobre ciertas clases de señales de pulso. La elección de una definición particular para el ancho de banda cambia el valor de la constante.

Para ser más específicos, considere el ancho de banda definido en la ecuación (A2.25). La definición correspondiente para la *duración rms* de la señal $g(t)$ es

$$T_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 \, dt}{\int_{-\infty}^{\infty} |g(t)|^2 \, dt} \right)^{1/2} \tag{A2.26}$$

donde se supone que la señal $g(t)$ se centra alrededor del origen. Puede demostrarse que, utilizando las definiciones rms de las ecuaciones (A2.25) y (A2.26), que el producto tiempo-ancho de banda tiene la siguiente forma:

$$T_{\text{rms}} W_{\text{rms}} \geq \frac{1}{4\pi} \tag{A2.27}$$

donde la constante es $1/4\pi$. El pulso gaussiano $\exp(-\pi t^2)$ satisface esta situación con el signo de igualdad.

■ ANCHO DE BANDA EQUIVALENTE DEL RUIDO

Las definiciones del ancho de banda que acaban de presentarse (es decir, el ancho de banda de 3 dB, el ancho de banda de nulo a nulo y el ancho de banda rms) se formulan en términos de señales determinísticas. Otra definición del ancho de banda que se presenta por sí sola en el estudio de las señales y los sistemas aleatorios es el ancho de banda equivalente del ruido. Suponga que una fuente de ruido blanco de densidad espectral de potencia $N_0/2$ se conecta a la entrada del filtro pasobajas RC simple de la figura A2.1; el valor correspondiente de la potencia promedio del ruido de salida es igual a $N_0/(4RC)$. Para este filtro, el ancho de banda de potencia media o de 3 dB es igual a $1/(2\pi RC)$. En este caso encontramos de nuevo que la potencia promedio del ruido de salida del filtro es proporcional al ancho de banda.

Es posible generalizar este enunciado para incluir todos los tipos de filtros pasobajas definiendo un ancho de banda equivalente de ruido de la manera siguiente. Suponga que tenemos una fuente de ruido blanco de media cero y densidad espectral de potencia $N_0/2$ conectada a la

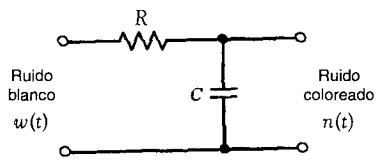


FIGURA A2.1 Filtro pasobajas RC.

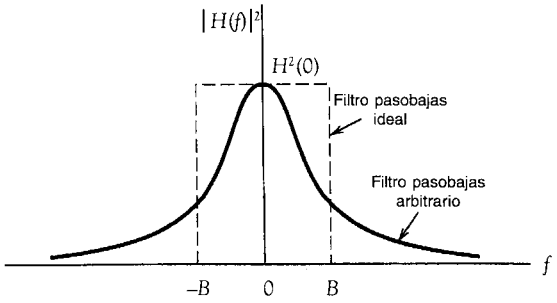


FIGURA A2.2 Ilustración de la definición del ancho de banda equivalente del ruido para un filtro pasobajas.

entrada de un filtro pasobajas arbitrario de función de transferencia $H(f)$. La potencia promedio del ruido de salida que resulta es entonces

$$\begin{aligned} N_{\text{salida}} &= \frac{N_0}{2}\int_{-\infty}^{\infty}|H(f)|^2\,df \\ &= N_0\int_0^{\infty}|H(f)|^2\,df \end{aligned} \tag{A2.28}$$

donde, en la última línea hemos usado el hecho de que la respuesta en magnitud $|H(f)|$ es una función par de la frecuencia.

Considere a continuación la misma fuente de ruido blanco conectada a la entrada de un filtro pasobajas *ideal* de respuesta en frecuencia cero $H(0)$ y ancho de banda B. En este caso, la potencia promedio del ruido de salida es

$$N_{\text{salida}} = N_0 B H^2(0) \tag{A2.29}$$

Por tanto, igualando esta potencia promedio del ruido de salida con la de la ecuación (A2.28), podemos definir formalmente el *ancho de banda equivalente del ruido* como

$$B = \frac{\int_0^{\infty}|H(f)|^2\,df}{H^2(0)} \tag{A2.30}$$

De este modo el procedimiento para calcular el ancho de banda equivalente del ruido consiste en sustituir el filtro pasobajas arbitrario de la función de transferencia $H(f)$ por un filtro pasobajas ideal equivalente de respuesta en frecuencia cero $H(0)$ y ancho de banda B, como se ilustra en la figura A2.2. De un modo similar, es posible definir un ancho de banda equivalente en el ruido para los filtros pasobanda.

A2.3 *Transformada de Hilbert*

La transformada de Fourier es particularmente útil para evaluar la frecuencia de una señal de energía o, en un sentido limitado, el de la señal de potencia. Como tal, ofrece las bases matemáticas para analizar y diseñar filtros selectivos de frecuencia para la separación de señales con base en su contenido de frecuencia. Otro método de separar señales se basa en la *selectividad de fase*, que recurre a los desfasamientos entre las señales pertinentes para alcanzar la separación deseada. El corrimiento de fase más simple es el de 180 grados, el cual es una inversión de polaridad en el caso de una señal senoidal. El corrimiento de los ángulos de fase de todas las componentes de una señal dada en 180 grados requiere el uso de un *transformador ideal*. Otro corrimiento de fase de interés es el correspondiente a ±90 grados. En particular, cuando los

ángulos de fase de todas las componentes de una señal dada se recorren en ±90 grados, la función resultante de tiempo se conoce como la transformada de Hilbert de la señal.

Para ser específicos, considere una señal $g(t)$ con transformada de Fourier $G(f)$. La *transformada de Hilbert* de $g(t)$, la cual denotaremos mediante $\hat{g}(t)$, está definida por

$$\hat{g}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau}\, d\tau \tag{A2.31}$$

Claramente, la transformación de Hilbert de $g(t)$ es una operación lineal. La *transformada inversa de Hilbert*, por medio de la cual la señal $g(t)$ se recupera a partir de $\hat{g}(t)$, se define mediante

$$g(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\hat{g}(\tau)}{t-\tau}\, d\tau \tag{A2.32}$$

Se dice que las funciones $g(t)$ y $\hat{g}(t)$ constituyen un *par de transformadas de Hilbert*. Una representación breve de pares de transformadas de Hilbert se presenta en la tabla A6.4.

Advierta, a partir de la definición de la transformada de Hilbert, que $\hat{g}(t)$ puede interpretarse como la convolución de $g(t)$ con la función de tiempo $1/\pi t$. Sabemos también, a partir del teorema de convolución, que la convolución de dos funciones en el dominio del tiempo se transforma en la multiplicación de sus transformadas de Fourier en el dominio de la frecuencia; vea la entrada 12 de la tabla A6.2. Para la función de tiempo $1/\pi t$, tenemos (vea la tabla A6.3)

$$\frac{1}{\pi t} \rightleftharpoons -j\,\mathrm{sgn}(f) \tag{A2.33}$$

donde $\mathrm{sgn}(f)$ es la *función signo* definida en el dominio de la frecuencia como

$$\mathrm{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \tag{A2.34}$$

Por tanto, se concluye que la transformada de Fourier $\hat{G}(f)$ de $\hat{g}(t)$ está dada por

$$\hat{G}(f) = -j\ \mathrm{sgn}(f)G(f) \tag{A2.35}$$

La ecuación (A2.35) establece que dada una señal $g(t)$, es posible obtener su transformada de Hilbert $\hat{g}(f)$ haciendo pasar $g(t)$ a través de un dispositivo lineal de dos puertos cuya respuesta en frecuencia sea igual a $-j\,\mathrm{sgn}(f)$. Este dispositivo puede considerarse como uno que produce un desfasamiento de −90 grados para todas las frecuencias positivas de la señal de entrada y de +90 grados para todas las frecuencias negativas, como en la figura A2.3. Sin embargo, las amplitudes

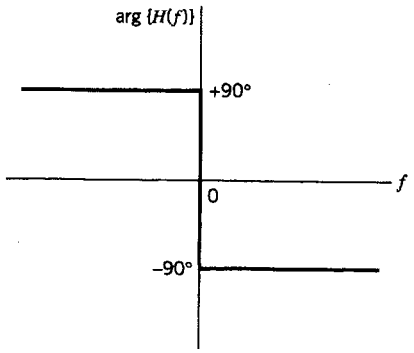


FIGURA A2.3 Característica de fase de un dispositivo lineal de dos puertos para obtener la transformada de Hilbert de una señal de valores reales.

de todas las componentes de frecuencia en la señal no son afectadas por la transmisión a través del dispositivo. Un dispositivo ideal de estas características se conoce como *transformador de Hilbert.*

Propiedades de la transformada de Hilbert

La transformada de Hilbert difiere de la transformada de Fourier porque opera exclusivamente en el dominio del tiempo. Tiene varias propiedades útiles, algunas de las cuales se indican a continuación. La señal $g(t)$ se supone de valores reales, lo cual constituye el dominio de aplicación usual de la transformada de Hilbert. Para esta clase de señales, es posible establecer lo siguiente:

1. Una señal $g(t)$ y su transformada de Hilbert $\hat{g}(t)$ tienen el mismo espectro de magnitud.
2. Si $\hat{g}(t)$ es la transformada de Hilbert de $g(t)$, entonces la transformada de Hilbert de $\hat{g}(t)$ es $-g(t)$.
3. Una señal $g(t)$ y su transformada de Hilbert $\hat{g}(t)$ son ortogonales sobre el intervalo de tiempo completo $(-\infty, \infty)$, según indica

$$\int_{-\infty}^{\infty} g(t)\hat{g}(t)dt = 0$$

Las demostraciones de estas propiedades se dejan como ejercicios al lector; éstas siguen de las ecuaciones (A2.31), (A2.32) y (A2.35).

A2.4 *Representación compleja de señales y sistemas*

Preenvolvente

Considere una señal de valores reales $g(t)$. Definimos la *preenvolvente* o *señal analítica* de la señal $g(t)$ como la función de valores complejos

$$g_+(t) = g(t) + j\hat{g}(t) \tag{A2.36}$$

donde $\hat{g}(t)$ es la transformada de Hilbert de $g(t)$. Observamos que la señal dada $g(t)$ es la parte real de la preenvolvente $g_+(t)$, y que la transformada de Hilbert de la señal es la parte imaginaria de la preenvolvente. Del mismo modo que el uso de fasores simplifica las manipulaciones de corrientes y voltajes alternos, así encontramos que la preenvolvente resulta particularmente útil en el manejo de señales y sistemas pasobanda.

Una de las características importantes de la preenvolvente $g_+(t)$ es el comportamiento de su transformada de Fourier. Sea $G_+(f)$ la transformada de Fourier de $g_+(t)$. Entonces podemos escribir

$$G_+(f) = G(f) + \text{sgn}(f)G(f)$$

a partir de lo cual encontramos fácilmente que

$$G_+(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} \tag{A2.37}$$

donde $G(0)$ es el valor de $G(f)$ a la frecuencia $f = 0$. Esto significa que la preenvolvente de una señal no tiene contenido de frecuencia (es decir, su transformada de Fourier se anula) para todas las frecuencias negativas.

A partir del análisis anterior es claro que para una señal dada $g(t)$ es posible determinar su preenvolvente $g_+(t)$ en una de las dos formas equivalentes:

1. Determinamos la transformada de Hilbert $\hat{g}(t)$ de la señal $g(t)$, y utilizamos después la ecuación (A2.36) para calcular la preenvolvente $g_+(t)$.
2. Determinamos la transformada de Fourier $G(f)$ de la señal $g(t)$, utilizamos la ecuación (A2.37) para determinar $G_+(f)$ y evaluamos después la transformada inversa de Fourier $G_+(f)$ para obtener

$$g_+(t) = 2\int_0^\infty G(f)\ \exp(j2\pi ft)\, df \tag{A2.38}$$

Para una señal particular $g(t)$ de transformada de Fourier $G(f)$, una de estas formas quizá sea mejor que la otra.

La ecuación (A2.36) define la preenvolvente $g_+(t)$ para frecuencias positivas. Simétricamente, es posible definir la preenvolvente para *frecuencias negativas* como

$$g_-(t) = g(t) - j\hat{g}(t) \tag{A2.39}$$

Las dos preenvolventes $g_+(t)$ y $g_-(t)$ son simplemente las complejas conjugadas mutuas, como indica

$$g_-(t) = g_+^*(t) \tag{A2.40}$$

donde el asterisco denota conjugación compleja. El espectro de la preenvolvente $g_+(t)$ es distinto de cero sólo para frecuencias *positivas*, como se subraya en la ecuación (A2.37); de ahí el uso del signo más como subíndice. En contraste, el espectro de la otra preenvolvente $g_-(t)$ es distinto de cero sólo para frecuencias *negativas*, como se indica mediante la transformada de Fourier

$$G_-(f) = \begin{cases} 0, & f > 0 \\ G(0), & f = 0 \\ 2G(f), & f < 0 \end{cases} \tag{A2.41}$$

De este modo las preenvolventes $g_+(t)$ y $g_-(t)$ constituyen un par complementario de señales de valor complejo. Advierta también que la suma de $g_+(t)$ y $g_-(t)$ es exactamente el doble de la señal original $g(t)$.

■ REPRESENTACIONES CANÓNICAS DE LAS SEÑALES PASOBANDA

Considere una señal pasobanda $g(t)$ cuya transformada de Fourier $G(f)$ no es ignorable únicamente en una banda de frecuencias de extensión total $2W$, digamos, centrada alrededor de alguna frecuencia $\pm f_c$, lo cual se ilustra en la figura A2.4*a*. Nos referimos a f_c como la *frecuencia de la portadora*. En la mayoría de las señales de comunicación, encontramos que el ancho de banda $2W$ es pequeño en comparación con f_c, por lo que nos referimos a ellas como *señales de banda angosta*. Sin embargo, un enunciado preciso acerca de qué tan pequeño debe ser el ancho de banda para que la señal sea considerada de banda angosta, no es necesario para nuestra explicación presente.

Considere que la preenvolvente de una señal de banda angosta $g(t)$, con su transformada de Fourier $G(f)$ centrada en torno a alguna frecuencia $\pm f_c$, se expresa en la forma

$$g_+(t) = \tilde{g}(t)\ \exp(j2\pi f_c t) \tag{A2.42}$$

Nos referimos a $\tilde{g}(t)$ como la *envolvente compleja* de la señal. La ecuación (A2.42) puede considerarse como la base de una definición para la envolvente compleja $\tilde{g}(t)$ en términos de la preenvolvente $g_+(t)$. Advertimos que el espectro de $g_+(t)$ está limitado a la banda de frecuencia $f_c - W \le f \le f_c + W$, como se ilustra en la figura A2.4*b*. Por tanto, al aplicar la propiedad del corrimiento de frecuencia de la transformada de Fourier a la ecuación (A2.42), la cual se des-

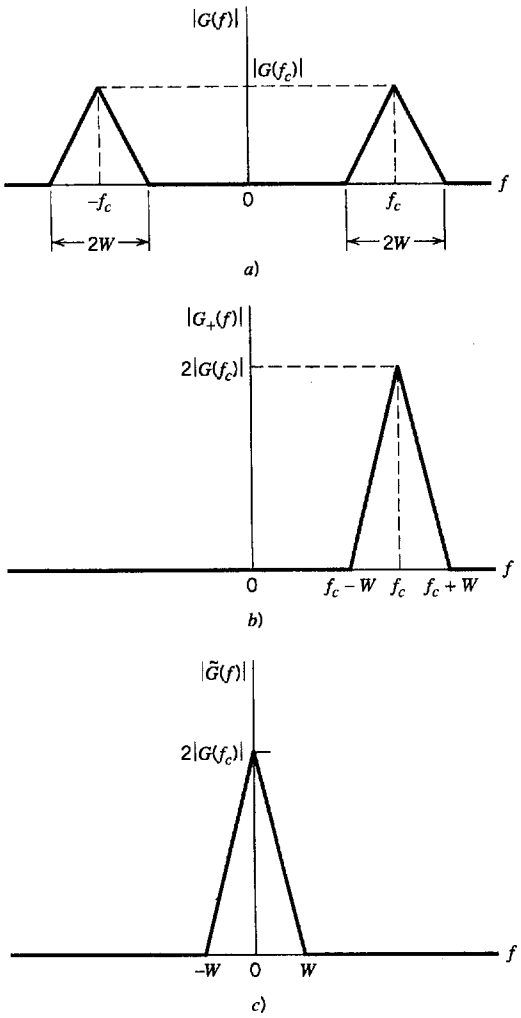


FIGURA A2.4 *a*) Espectro de magnitud de la señal pasobanda $g(t)$. *b*) Espectro de magnitud de la preenvolvente $g_+(t)$. *c*) Espectro de magnitud de la envolvente compleja $\tilde{g}(t)$.

cribe como la entrada 5 en la tabla A6.2, encontramos que el espectro de la envolvente compleja $\tilde{g}(t)$ está limitado a la banda $-W \leq f \leq W$ y centrado en el origen, como se ilustra en la figura A2.4c. Es decir, la envolvente compleja $\tilde{g}(t)$ de una señal pasobanda $g(t)$ es una *señal pasobajas*, lo que constituye un resultado importante.

Por definición, la señal dada $g(t)$ es la parte real de la preenvolvente $g_+(t)$. Podemos entonces expresar la señal pasobajas original $g(t)$ en términos de la envolvente compleja $\tilde{g}(t)$ como sigue:

$$g(t) = \text{Re}[\tilde{g}(t) \exp(j2\pi f_c t)] \tag{A2.43}$$

En general, $\tilde{g}(t)$ es una cantidad de valores complejos; para subrayar esta propiedad, es posible expresarla

$$\tilde{g}(t) = g_I(t) + jg_Q(t) \tag{A2.44}$$

donde $g_I(t)$ y $g_Q(t)$ son funciones pasobajas de valor real. Su propiedad pasobajas es heredada de la envolvente compleja $\tilde{g}(t)$. Por lo tanto, podemos utilizar la ecuación (A2.43) y la (A2.44) para expresar la señal pasobanda original $g(t)$ en la forma canónica o estándar:

$$g(t) = g_I(t)\ \cos(2\pi f_c t) - g_Q(t)\ \text{sen}(2\pi f_c t) \tag{A2.45}$$

Nos referimos a $g_I(t)$ como la componente en fase de la señal pasobanda $g(t)$ y a $g_Q(t)$ como la componente en cuadratura de la señal; esta nomenclatura admite que $\text{sen}(2\pi f_c t)$ [es decir, el factor multiplicador de $g_Q(t)$] está en fase y en cuadratura con respecto a $\cos(2\pi f_c t)$ [es decir, el factor multiplicador de $g_I(t)$] y $\cos(2\pi f_c t)$ se considera como la referencia.

De acuerdo con la ecuación (A2.44), la envolvente compleja $\tilde{g}(t)$ puede imaginarse como un *fasor invariante con el tiempo* ubicado en el origen del plano (g_I, g_Q), como se indica en la figura A2.5*a*. Con el tiempo t variable, el extremo del fasor se mueve en torno al plano. La figura A2.5*b* muestra la representación fasorial de la exponencial compleja $\exp(j2\pi f_c t)$. En la definición dada en la ecuación (A2.43), la envolvente compleja $\tilde{g}(t)$ se multiplica por la exponencial compleja $\exp(j2\pi f_c t)$. En consecuencia, los ángulos de estos dos fasores se suman y sus longitudes se multiplican, como se muestra en la figura A2.5*c*. Además, en esta última figura ilustramos el plano (g_I, g_Q) girando con una velocidad angular igual a $2\pi f_c$ radianes por segundo. De ese modo, en la imagen que se ilustra aquí, el fasor que representa la envolvente compleja $\tilde{g}(t)$ se mueve en el plano (g_I, g_Q) y al mismo tiempo el propio plano gira alrededor del origen. La señal pasobanda original $g(t)$ es la proyección de este fasor variante con el tiempo sobre una *línea fija* que representa al eje real, como se indica en la figura A2.5c.

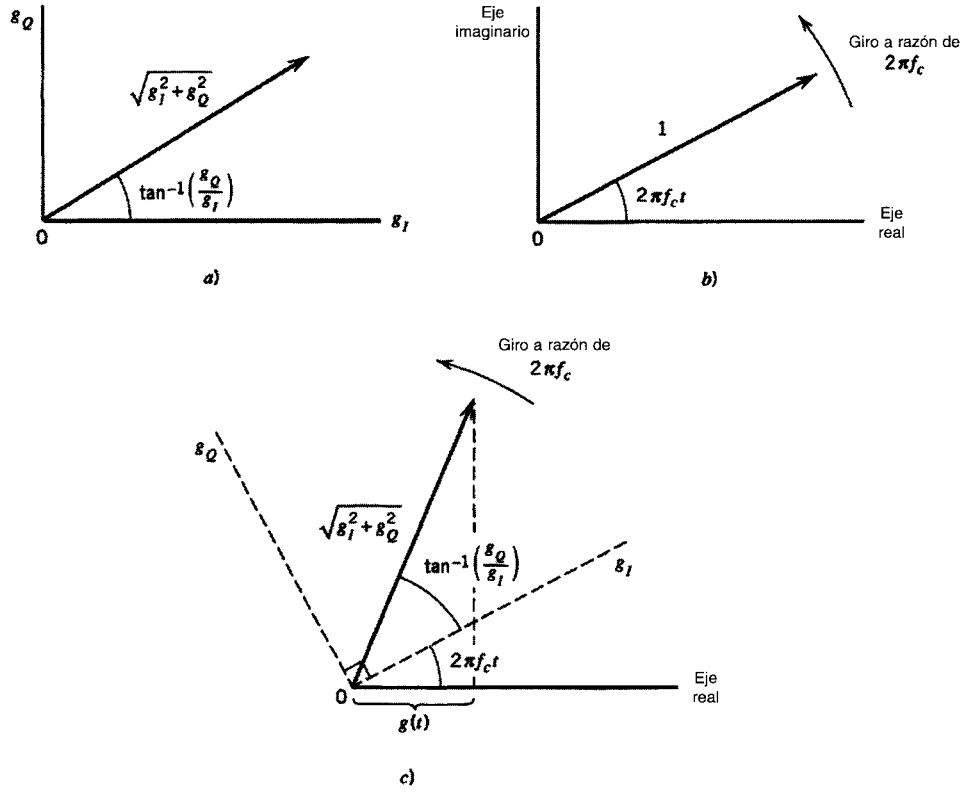


FIGURA A2.5 Ilustración de una interpretación de la envolvente compleja $\tilde{g}(t)$ y de su multiplicación por $\exp(j2\pi f_c t)$.

Puesto que tanto $g_I(t)$ como $g_Q(t)$ son señales pasobajas limitadas a la banda $-W \leq f \leq W$, éstas pueden obtenerse a partir de la señal pasobanda $g(t)$ utilizando el esquema que se muestra en la figura A2.6*a*. Ambos filtros pasobajas en esta figura son idénticos, y cada uno tiene un ancho de banda igual a W. Para reconstruir $g(t)$ a partir de sus componentes en fase y cuadratura, es posible utilizar el esquema que se presenta en la figura A2.6*b*.

Los dos esquemas de la figura A2.6 resultan básicos para el estudio de los *sistemas de modulación lineales*. La multiplicación de la componente en fase pasobajas $g_I(t)$ por $\cos(2\pi f_c t)$ y la multiplicación de la componente pasobajas en cuadratura $g_Q(t)$ por $\text{sen}(2\pi f_c t)$ representan formas lineales de modulación. Dado que la frecuencia f_c de la portadora es suficientemente grande, la función pasobanda resultante $g(t)$ definida en la ecuación (A2.45) se conoce como una *forma de onda de transmisión de señales pasobanda*. De modo paralelo, la correspondencia de $g_I(t)$ y $g_Q(t)$ con $g(t)$ se conoce como *modulación pasobanda*.

La ecuación (A2.44) es la forma cartesiana de expresar la envolvente compleja $\tilde{g}(t)$. De manera alternativa, es posible expresarla en la forma polar

$$\tilde{g}(t) = a(t)\exp[j\phi(t)] \tag{A2.46}$$

donde $a(t)$ y $\phi(t)$ son funciones pasobajas de valor real. Con base en esta representación polar, la señal pasobanda original $g(t)$ se define por medio de

$$g(t) = a(t)\cos[2\pi f_c t + \phi(t)] \tag{A2.47}$$

Nos referimos a $a(t)$ como la *envolvente natural* o simplemente la *envolvente* de la señal pasobanda $g(t)$ y a $\phi(t)$ como a la *fase* de la señal. La ecuación (A2.47) representa una *forma híbrida de modulación de amplitud y modulación angular*; en realidad, incluye la modulación de amplitud, la modulación de frecuencia y la modulación de fase como casos especiales.

A partir de esta explicación es claro que, si representamos una señal pasobanda (modulada) $g(t)$ en términos de sus componentes en fase y cuadratura, como en la ecuación (A2.45), o en términos de su envolvente y fase, como en la ecuación (A2.47), el contenido de información de la señal $g(t)$ se preserva completamente en la envolvente compleja $\tilde{g}(t)$.

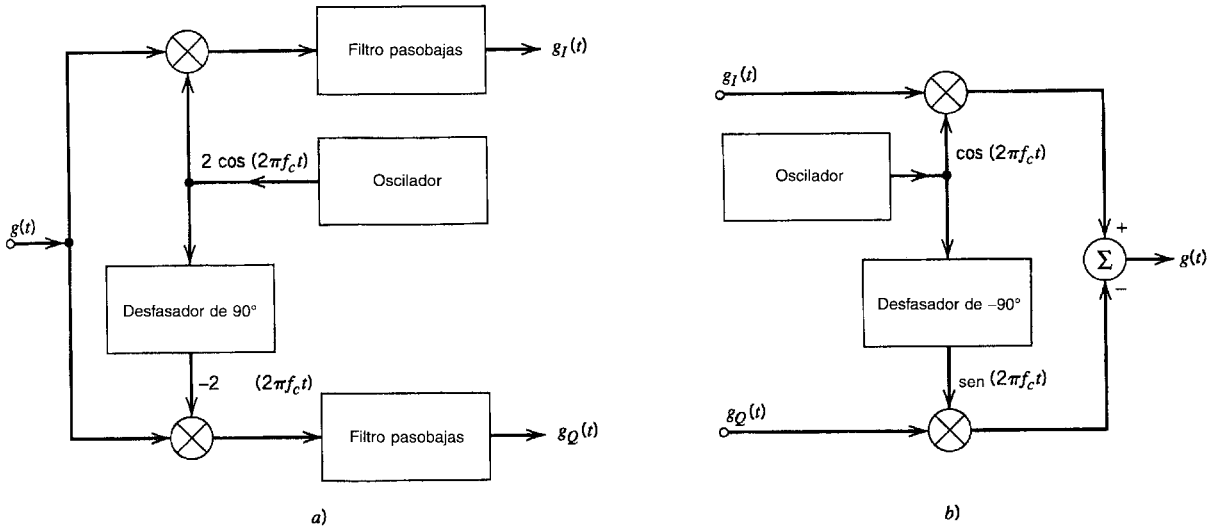


FIGURA A2.6 *a*) Esquema para derivar las componentes en fase y en cuadratura de una señal pasobanda. *b*) Esquema para reconstruir la señal pasobanda a partir de sus componentes en fase y en cuadratura.

■ TERMINOLOGÍA

Las distinciones entre las tres diferentes envolventes que hemos presentado para describir una señal pasobanda $g(t)$ deben advertirse con todo cuidado. A continuación resumimos sus definiciones:

1. La preenvolvente $g_+(t)$ para frecuencias positivas está definida por

$$g_+(t) = g(t) + j\hat{g}(t)$$

donde $\hat{g}(t)$ es la transformada de Hilbert de la señal $g(t)$. De acuerdo con esta representación, $\hat{g}(t)$ puede considerarse como la función en cuadratura de $g(t)$. De modo correspondiente, en el dominio de la frecuencia tenemos

$$G_+(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases}$$

2. La envolvente compleja $\tilde{g}(t)$ es igual a una versión recorrida en frecuencia de la preenvolvente $g_+(t)$, como se indica mediante

$$\tilde{g}(t) = g_+(t) \exp(-j2\pi f_c t)$$

donde f_c es la frecuencia de la portadora de la señal pasobanda $g(t)$.

3. La envolvente $a(t)$ es igual a la magnitud de la envolvente compleja $\tilde{g}(t)$ y también a la de la preenvolvente $g_+(t)$, según indica

$$a(t) = |\tilde{g}(t)| = |g_+(t)|$$

Advierta que para una señal pasobanda $g(t)$, la preenvolvente $g_+(t)$ es una señal pasobanda compleja cuyo valor depende de la frecuencia f_c de la portadora. Por otro lado, la envolvente $a(t)$ es siempre una señal pasobajas real y, en general, la envolvente compleja $\tilde{g}(t)$ es una señal pasobajas compleja; los valores de las últimas dos envolventes son independientes de la elección de la frecuencia f_c de la portadora. Esta propiedad proporciona a la envolvente compleja $\tilde{g}(t)$ una ventaja analítica sobre la señal $g(t)$.

La envolvente $a(t)$ y la fase $\phi(t)$ de $g(t)$ se relacionan con las componentes en fase y en cuadratura $g_I(t)$ y $g_Q(t)$ como sigue (vea la representación fasorial variable en el tiempo de la figura A2.5*a*):

$$a(t) = \sqrt{g_I^2(t) + g_Q^2(t)}$$

$$\phi(t) = \tan^{-1}\left(\frac{g_Q(t)}{g_I(t)}\right)$$

Inversamente, es posible escribir

$$g_I(t) = a(t)\ \cos[\phi(t)]$$

$$g_Q(t) = a(t)\ \text{sen}[\phi(t)]$$

De este modo, cada una de las componentes en cuadratura de la señal pasobanda contiene información tanto de amplitud como de fase. Ambas componentes se requieren para una definición única de la fase $\phi(t)$, módulo 2π.

■ SISTEMAS PASOBANDA

Ahora que sabemos cómo manejar la representación pasobajas compleja de las señales pasobanda, resulta lógico que formulemos un procedimiento correspondiente para manejar el análisis de los

sistemas pasobanda. Específicamente, deseamos demostrar que el análisis de los sistemas pasobanda puede simplificarse en forma considerable estableciendo una analogía (o, más precisamente, un isomorfismo) entre los sistemas pasobajas y pasobanda. Esta analogía se basa en el uso de la transformada de Hilbert para la representación de señales pasobanda.

Considere una señal de banda angosta $x(t)$, con su transformada de Fourier $X(f)$. Suponemos que el espectro de la señal $x(t)$ está limitado a frecuencias de $\pm W$ Hz de la frecuencia f_c de la portadora. Además, suponemos que $W < f_c$. Sea esta señal representada en términos de sus componentes en fase y en cuadratura de la manera siguiente:

$$x(t) = x_I(t)\ \cos(2\pi f_c t) - x_Q(t)\ \operatorname{sen}(2\pi f_c t) \tag{A2.48}$$

donde $x_I(t)$ es la componente en fase y $x_Q(t)$ es la componente en cuadratura. En ese caso, utilizando $x(t)$ para denotar la envolvente compleja de $x(t)$, podemos expresar

$$\tilde{x}(t) = x_I(t) + jx_Q(t) \tag{A2.49}$$

Considere que la señal $x(t)$ se aplica en un sistema pasobanda lineal invariante con el tiempo con respuesta al impulso $h(t)$ y respuesta en frecuencia $H(f)$. Suponemos que la respuesta en frecuencia del sistema está limitada a frecuencias dentro de $\pm B$ de la frecuencia f_c de la portadora. El ancho de banda del sistema $2B$ suele ser más estrecho o igual que el ancho de banda de la señal de entrada $2W$. Deseamos representar la respuesta al impulso pasobanda $h(t)$ en términos de las dos componentes en cuadratura denotadas por $h_I(t)$ y $h_Q(t)$. De ese modo, mediante la analogía para la representación de señales pasobanda, es posible expresar $h(t)$ en la forma

$$h(t) = h_I(t)\ \cos(2\pi f_c t) - h_Q(t)\ \operatorname{sen}(2\pi f_c t) \tag{A2.50}$$

Defina la *respuesta al impulso compleja* del sistema pasobanda como

$$\tilde{h}(t) = h_I(t) + jh_Q(t) \tag{A2.51}$$

Por tanto, tenemos la representación compleja

$$h(t) = \operatorname{Re}[\tilde{h}(t)\ \exp(j2\pi f_c t)] \tag{A2.52}$$

Observe que $h_I(t)$, $h_Q(t)$ y $\tilde{h}(t)$ son funciones pasobajas limitadas a la banda de frecuencia $-B \le f \le B$.

Podemos determinar la respuesta al impulso compleja $\tilde{h}(t)$ en términos de las componentes en cuadratura $h_I(t)$ y $h_Q(t)$ de la respuesta al impulso pasobanda $h(t)$ utilizando la ecuación (A2.51). De modo alternativo, se determina a partir de la respuesta en frecuencia pasobanda $H(f)$ de la forma siguiente. Notamos primero, de acuerdo con la ecuación (A2.52), que

$$2h(t) = \tilde{h}(t)\exp(j2\pi f_c t) + \tilde{h}^*(t)\ \exp(-j2\pi f_c t) \tag{A2.53}$$

donde $\tilde{h}^*(t)$ es la compleja conjugada de $\tilde{h}(t)$. Por tanto, aplicando la transformada de Fourier a la ecuación (A2.53) y utilizando la propiedad de la conjugación compleja de la transformada de Fourier, la cual se describe en la entrada 10 de la tabla A6.2, obtenemos

$$2H(f) = \tilde{H}(f - f_c) + \tilde{H}^*(-f - f_c) \tag{A2.54}$$

donde $H(f)$ es la transformada de Fourier de $h(t)$ y $\tilde{H}(f)$ es la transformada de Fourier de $\tilde{h}(t)$. La ecuación (A2.54) satisface el requisito de que $H^*(f) = H(-f)$ para una respuesta al impulso real $h(t)$. Puesto que $\tilde{H}(f)$ representa una respuesta en frecuencia pasobajas limitada a $|f| \le B$ con $B < f_c$, deducimos a partir de la ecuación (A2.54) que

$$\tilde{H}(f - f_c) = 2H(f), \qquad f > 0 \tag{A2.55}$$

La ecuación (A2.55) indica que para una respuesta en frecuencia pasobanda especificada $H(f)$, podemos determinar $\widetilde{H}(f)$ tomando la parte de $H(f)$ correspondiente a las frecuencias positivas, recorriéndola hacia el origen y escalándola después por el factor 2. Para determinar la respuesta al impulso compleja $h(t)$, tomamos la transformada inversa de Fourier $\widetilde{H}(f)$, obteniendo

$$\widetilde{h}(t) = \int_{-\infty}^{\infty} \widetilde{H}(f) \exp(j2\pi ft)\, df \tag{A2.56}$$

Las representaciones que acaban de escribirse para las señales y los sistemas pasobanda proporcionan las bases de un método eficiente de determinación de la salida de un sistema pasobanda accionado por una señal pasobanda. Suponemos que el espectro de la señal de entrada $x(t)$ y la respuesta en frecuencia $H(f)$ del sistema están centradas alrededor de la frecuencia f_c. En la práctica no hay necesidad de considerar una situación en la cual la frecuencia de la portadora de la señal de entrada no está alineada con la frecuencia de banda media del sistema pasobanda, puesto que tenemos una libertad considerable de elegir la frecuencia de la portadora o de la banda media. De ese modo, cambiar la frecuencia de la portadora de la señal de entrada por una cantidad Δf_c, digamos, corresponde simplemente a absorber (o eliminar) el factor $\exp(\pm j2\pi \Delta f_c t)$ en la envolvente compleja de la señal de entrada o la respuesta al impulso compleja del sistema pasobanda. Por tanto, tenemos justificación para proceder con la suposición de que $X(f)$ y $H(f)$ están centradas alrededor de f_c. Supóngase entonces que utilizamos $y(t)$ para denotar la salida del sistema. Es claro que $y(t)$ es también una señal pasobanda, por lo que es posible representarla en términos de su envolvente compleja pasobajas $\widetilde{y}(t)$, de la manera siguiente:

$$y(t) = \mathrm{Re}[\widetilde{y}(t)\ \exp(j2\pi f_c t)] \tag{A2.57}$$

La señal de salida $y(t)$ se relaciona con la señal $x(t)$ y con la respuesta de impulso $h(t)$ del sistema en la forma usual mediante la integral de convolución

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\, d\tau \tag{A2.58}$$

En términos de las preenvolventes, tenemos que $h(t) = \mathrm{Re}[h_+(t)]$ y $x(t) = \mathrm{Re}[x_+(t)]$. Por tanto, es posible escribir la ecuación (A2.58) en términos de las preenvolventes $x_+(t)$ y $h_+(t)$ como sigue:

$$y(t) = \int_{-\infty}^{\infty} \mathrm{Re}[h_+(\tau)]\ \mathrm{Re}[x_+(t-\tau)]\, d\tau \tag{A2.59}$$

Para continuar, hacemos uso de la propiedad básica de las preenvolventes que se describe mediante la siguiente relación (presentada aquí sin demostración):

$$\int_{-\infty}^{\infty} \mathrm{Re}[h_+(\tau)]\ \mathrm{Re}[x_+(\tau)]\, d\tau = \frac{1}{2}\mathrm{Re}\left[\int_{-\infty}^{\infty} h_+(\tau)x_+^*(\tau)\, d\tau\right] \tag{A2.60}$$

donde hemos usado τ como la variable de integración para ser consistentes con la de la ecuación (A2.59). Posteriormente observamos que el uso de $x(-\tau)$ en lugar de $x(\tau)$ tiene el efecto de eliminar la conjugación compleja en el segundo miembro de la ecuación (A2.60). De ese modo, teniendo en mente la diferencia algebraica entre el argumento de $x_+(\tau)$ en la ecuación (A2.60) y la correspondiente a $x_+(t-\tau)$ en la ecuación (A2.59), y utilizando la relación entre la preenvolvente y la envolvente compleja de una función pasobanda, obtenemos

$$\begin{aligned} y(t) &= \frac{1}{2}\mathrm{Re}\left[\int_{-\infty}^{\infty} h_+(\tau)x_+(t-\tau)\, d\tau\right] \\ &= \frac{1}{2}\mathrm{Re}\left[\int_{-\infty}^{\infty} \widetilde{h}(\tau)\exp(j2\pi f_c\tau)\widetilde{x}(t-\tau)\ \exp(j2\pi f_c(t-\tau))\, d\tau\right] \\ &= \frac{1}{2}\mathrm{Re}\left[\exp(j2\pi f_c t)\int_{-\infty}^{\infty} \widetilde{h}(\tau)\widetilde{x}(t-\tau)\, d\tau\right] \end{aligned} \tag{A2.61}$$

Comparando de ese modo el segundo miembro de las ecuaciones (A2.57) y (A2.61), deducimos sin dificultad que para una frecuencia f_c de portadora suficientemente grande, la envolvente compleja $\tilde{y}(t)$ de la señal de salida se relaciona con la envolvente compleja $\tilde{x}(t)$ de la señal de entrada y la respuesta al impulso compleja $\tilde{h}(t)$ del sistema pasobanda en la siguiente forma:

$$2\tilde{y}(t) = \int_{-\infty}^{\infty} \tilde{h}(\tau)\tilde{x}(t-\tau)\, d\tau \tag{A2.62}$$

o, utilizando la notación abreviada para la convolución,

$$2\tilde{y}(t) = \tilde{h}(t) \bigstar \tilde{x}(t) \tag{A2.63}$$

donde ★ denota convolución. En otras palabras, *excepto por el factor de escalamiento 2, la envolvente compleja $\tilde{y}(t)$ de la señal de salida de un sistema pasobanda se obtiene al aplicar la convolución a la respuesta al impulso compleja $\tilde{h}(t)$ del sistema con envolvente compleja $\tilde{x}(t)$ de la señal pasobanda de entrada.* La ecuación (A2.63) es el resultado del isomorfismo para la convolución entre la función pasobanda y la correspondiente función pasobajas.

La importancia de este resultado es que al tratar con señales y sistemas pasobanda, sólo necesitamos interesarnos en las funciones pasobajas $\tilde{x}(t)$, $\tilde{y}(t)$ y $\tilde{h}(t)$, que representan la excitación, la respuesta y el sistema, respectivamente. Es decir, el análisis de un sistema pasobanda, el cual se complica por la presencia del factor multiplicador $\exp(j2\pi f_c t)$, se sustituye por un análisis pasobajas equivalente aunque mucho más simple que retiene por completo la esencia del proceso de filtrado. Este procedimiento se ilustra esquemáticamente en la figura A2.7.

La envolvente compleja $\tilde{x}(t)$ de la señal pasobanda de entrada y la respuesta al impulso compleja $\tilde{h}(t)$ del sistema pasobanda se definen en términos de sus componentes respectivas en fase y en cuadratura mediante las ecuaciones (A2.49) y (A2.51), respectivamente. Al sustituir estas relaciones en la ecuación (A2.63), obtenemos

$$2\tilde{y}(t) = [h_I(t) + jh_Q(t)] \bigstar [x_I(t) + jx_Q(t)] \tag{A2.64}$$

Debido a que la convolución es *distributiva*, es posible reescribir la ecuación (A2.64) en la forma equivalente

$$2\tilde{y}(t) = [h_I(t) \bigstar x_I(t) - h_Q(t) \bigstar x_Q(t)] + j[h_Q(t) \bigstar x_I(t) + h_I \bigstar x_Q(t)] \tag{A2.65}$$

Considere que la envolvente compleja $\tilde{y}(t)$ de la respuesta se defina en términos de sus componentes en fase y en cuadratura como

$$\tilde{y}(t) = y_I(t) + jy_Q(t) \tag{A2.66}$$

Al comparar las partes real e imaginaria en las ecuaciones (A2.65) y (A2.66), tenemos para la componente en fase $y_I(t)$ la relación

$$2y_I(t) = h_I(t) \bigstar x_I(t) - h_Q(t) \bigstar x_Q(t) \tag{A2.67}$$

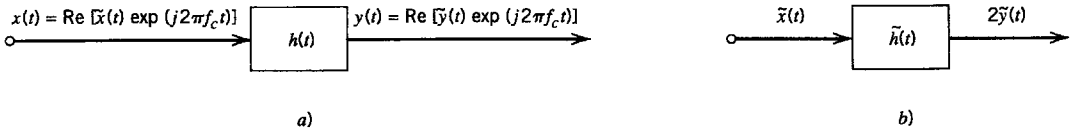


FIGURA A2.7 *a)* Filtro de banda angosta de la respuesta al impulso $h(t)$ con señal de entrada de banda angosta $x(t)$. *b)* Filtro pasobajas equivalente de respuesta al impulso compleja $\tilde{h}(t)$ con entrada pasobajas complejas $\tilde{x}(t)$.

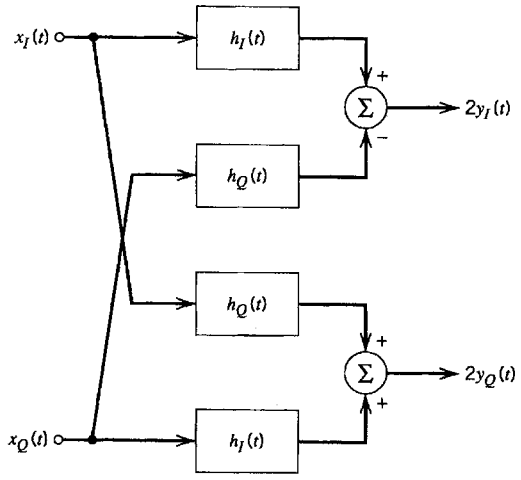


FIGURA A2.8 Diagrama de bloques que ilustra las relaciones entre las componentes en fase y en cuadratura de la respuesta de un filtro pasobanda y las correspondientes a la señal de entrada.

y para la componente en cuadratura $y_Q(t)$ la relación

$$2y_Q(t) = h_Q(t) \bigstar x_I(t) + h_I(t) \bigstar x_Q(t) \tag{A2.68}$$

Por tanto, con el fin de evaluar las componentes en fase y en cuadratura de la envolvente compleja $\tilde{y}(t)$ de la salida del sistema, podemos utilizar el *modelo equivalente pasobajas* ilustrado en la figura A2.8. Todas las señales y respuestas al impulso que se muestran en este modelo son funciones pasobajas de valores reales. Por tanto, este modelo equivalente ofrece una base práctica para la simulación eficiente de los filtros pasobanda o canales de comunicación en una computadora digital.

Resumiendo, el procedimiento para evaluar la respuesta de un sistema pasobanda (con frecuencia de banda media f_c) ante una señal pasobanda de entrada (de frecuencia f_c de portadora) es el siguiente:

1. La señal pasobanda de entrada $x(t)$ se sustituye por su envolvente compleja $\hat{x}(t)$, la cual se relaciona con $x(t)$ mediante

$$x(t) = \text{Re}[\hat{x}(t) \exp(j2\pi f_c t)]$$

2. El sistema pasobanda, con respuesta al impulso $h(t)$, se sustituye por un analógico pasobanda, el cual se caracteriza mediante una respuesta al impulso compleja $\tilde{h}(t)$ relacionada con $h(t)$ por medio de

$$h(t) = \text{Re}[\tilde{h}(t) \exp(j2\pi f_c t)]$$

3. La envolvente compleja $\tilde{y}(t)$ de la señal pasobanda de salida $y(t)$ se obtiene al aplicar la convolución $\tilde{h}(t)$ con $\tilde{x}(t)$, como indica

$$2\tilde{y}(t) = \tilde{h}(t) \bigstar \tilde{x}(t)$$

4. La salida deseada $y(t)$ se obtiene finalmente de la envolvente compleja $\tilde{y}(t)$ utilizando la relación

$$y(t) = \text{Re}[\tilde{y}(t) \exp(j2\pi f_c t)]$$

APÉNDICE 3

FUNCIONES DE BESSEL

A3.1 *Solución en serie de la ecuación de Bessel*

En su forma más básica, la *ecuación de Bessel de orden n* se escribe como

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+(x^2-n^2)y=0 \tag{A3.1}$$

la cual es una de las ecuaciones diferenciales de coeficientes variables más importantes.[1] Para cada n, una solución de esta ecuación se define mediante la serie de potencias

$$J_n(x)=\sum_{m=0}^{\infty}\frac{(-1)^m\left(\frac{1}{2}x\right)^{n+2m}}{m!(n+m)!} \tag{A3.2}$$

La función $J_n(x)$ recibe el nombre de *función de Bessel del primer tipo de orden n*. La ecuación (A3.1) tiene dos funciones coeficientes, a saber, $1/x$ y $(1-n^2/x^2)$. Por tanto, no tiene puntos singulares finitos excepto el origen. En consecuencia, se concluye que una expansión en serie de la ecuación (A3.2) converge para toda $x > 0$. La ecuación (A3.2) puede, de este modo, utilizarse para calcular numéricamente $J_n(x)$ para $n = 0, 1, 2, \ldots$. La tabla A6.5 presenta valores de $J_n(x)$ para diferentes órdenes n y x variable. Es interesante advertir que las gráficas de $J_0(x)$ y $J_1(x)$ se asemejan a las gráficas de cos x y sen x, respectivamente; vea las gráficas de la figura 2.23 en el capítulo 2.

La función $J_n(x)$ puede expresarse también en la forma de una integral como

$$J_n(x)=\frac{1}{\pi}\int_0^{\pi}\cos(x\ \operatorname{sen}\theta-n\theta)\,d\theta \tag{A3.3}$$

o, de manera equivalente,

$$J_n(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(jx\ \operatorname{sen}\theta-jn\theta)\,d\theta \tag{A3.4}$$

A3.2 *Propiedades de la función de Bessel*

La función de Bessel $J_n(x)$ tiene las siguientes propiedades:

1. $$J_n(x)=(-1)^nJ_{-n}(x) \tag{A3.5}$$

Para demostrar esta relación, sustituimos θ por $(\pi-\theta)$ en la ecuación (A3.3). En ese caso, advirtiendo que $\operatorname{sen}(\pi-\theta)=\operatorname{sen}\theta$, obtenemos

$$\begin{aligned}J_n(x)&=\frac{1}{\pi}\int_0^{\pi}\cos(x\operatorname{sen}\theta+n\theta-n\pi)\,d\theta\\&=\frac{1}{\pi}\int_0^{\pi}[\cos(n\pi)\cos(x\ \operatorname{sen}\theta+n\theta)+\operatorname{sen}(n\pi)\ \operatorname{sen}(x\ \operatorname{sen}\theta+n\theta)]\,d\theta\end{aligned}$$

Para valores enteros de n, tenemos

$$\cos(n\pi) = (-1)^n$$
$$\operatorname{sen}(n\pi) = 0$$

Por tanto,

$$J_n(x) = \frac{(-1)^n}{\pi}\int_0^{\pi} \cos(x \operatorname{sen}\theta + n\theta)\, d\theta \tag{A3.6}$$

De acuerdo con la ecuación (A3.3), encontramos que al sustituir n por $-n$:

$$J_{-n}(x) = \frac{1}{\pi}\int_0^{\pi} \cos(x \operatorname{sen}\theta + n\theta)\, d\theta \tag{A3.7}$$

El resultado deseado se desprende de inmediato de las ecuaciones (A3.6) y (A3.7).

2.
$$J_n(x) = (-1)^n J_n(-x) \tag{A3.8}$$

Esta relación se obtiene sustituyendo x por $-x$ en la ecuación (A3.3) y utilizando después la ecuación (A3.6).

3.
$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x} J_n(x) \tag{A3.9}$$

Esta *fórmula de recurrencia* resulta útil al construir tablas de coeficientes de Bessel; su derivación sigue de la serie de potencias de la ecuación (A3.2).

4. Para valores pequeños de x, tenemos

$$J_n(x) \simeq \frac{x^n}{2^n n!} \tag{A3.10}$$

Esta relación se obtiene reteniendo simplemente el primer término de la serie de potencias de la ecuación (A3.2) ignorando los términos de orden superior. Por tanto, cuando x es pequeña, tenemos

$$\begin{aligned} J_0(x) &\simeq 1 \\ J_1(x) &\simeq \frac{x}{2} \\ J_n(x) &\simeq 0 \text{ para } n > 1 \end{aligned} \tag{A3.11}$$

5. Para valores grandes de x, ocurre que

$$J_n(x) \simeq \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi}{4} - \frac{n\pi}{2}\right) \tag{A3.12}$$

Esto muestra que para valores grandes de x, la función de Bessel $J_n(x)$ se comporta como una onda seno con amplitud progresivamente decreciente.

6. Con x real y fija, $J_n(x)$ tiende a cero cuando el orden n se aproxima a infinito.

7.
$$\sum_{n=-\infty}^{\infty} J_n(x) \exp(jn\phi) = \exp(jx \operatorname{sen}\phi) \tag{A3.13}$$

Para demostrar esta propiedad, considere la suma $\sum_{n=-\infty}^{\infty} J_n(x) \exp(jn\phi)$ y utilice la fórmula de la ecuación (A3.4) para $J_n(x)$ a fin de obtener

$$\sum_{n=-\infty}^{\infty} J_n(x) \exp(jn\phi) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \exp(jn\phi) \int_{-\pi}^{\pi} \exp(jx \operatorname{sen}\theta - jn\theta)\, d\theta$$

Intercambiando el orden de la integración y la sumatoria:

$$\sum_{n=-\infty}^{\infty} J_n(x)\ \exp(jn\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \exp(jx\ \text{sen}\,\theta) \sum_{n=-\infty}^{\infty} \exp[jn(\phi-\theta)] \tag{A3.14}$$

Apelamos ahora a la siguiente relación de la teoría de la transformada de Fourier:

$$\delta(\phi) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \exp[jn(\phi)], \qquad -\pi \le \phi \le \pi \tag{A3.15}$$

donde $\delta(\phi)$ es una función delta. Por tanto, utilizando la ecuación (A3.15) en la (A3.14) y aplicando luego la propiedad de selección de la función delta, obtenemos

$$\begin{aligned}\sum_{n=-\infty}^{\infty} J_n(x)\ \exp(jn\phi) &= \int_{-\pi}^{\pi} \exp(jx\ \text{sen}\,\theta)\,\delta(\phi-\theta)\,d\theta \\ &= \exp(jx\ \text{sen}\,\phi)\end{aligned}$$

que es el resultado deseado.

8. $$\sum_{n=-\infty}^{\infty} J_n^2(x) = 1 \quad \text{para toda } x \tag{A3.16}$$

Para demostrar esta propiedad, procedemos de la manera siguiente. Observamos que $J_n(x)$ es real. Por tanto, multiplicando la ecuación (A3.4) por su propia conjugada compleja y sumando sobre todo los valores posibles de n, obtenemos

$$\sum_{n=-\infty}^{\infty} J_n^2(x) = \frac{1}{(2\pi)^2} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp(jx\,\text{sen}\,\theta - jn\theta - jx\,\text{sen}\,\phi + jn\phi)\ d\theta\ d\phi$$

Intercambiando el orden de la doble integración y de la sumatoria:

$$\begin{aligned}&\sum_{n=-\infty}^{\infty} J_n^2(x) = \\ &\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} d\theta\,d\phi\ \exp[jx(\text{sen}\,\theta - \text{sen}\,\phi)] \sum_{n=-\infty}^{\infty} \exp[jn(\phi-\theta)]\end{aligned} \tag{A3.17}$$

Utilizando la ecuación (A3.15) en la (A3.17) y aplicando después la propiedad de selección de la función delta, obtenemos finalmente

$$\sum_{n=-\infty}^{\infty} j_n^2(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta = 1$$

que es el resultado que se deseaba.

Muchas de estas propiedades de la función de Bessel $J_n(x)$ pueden ilustrarse también en términos numéricos refiriéndonos a la tabla A6.5.

A3.3 *Función de Bessel modificada*

La *función de Bessel modificada de orden n* se escribe como

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + n^2)y = 0 \tag{A3.18}$$

Con $j^2 = -1$, donde j es la raíz cuadrada de -1, podemos reescribir esta ecuación en la forma

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (j^2x^2 - n^2)y = 0$$

A partir de aquí resulta evidente que la ecuación (A3.18) no es sino la ecuación de Bessel, a saber, la ecuación (A3.1), con x en vez de jx. Sustituyendo de esta manera x por jx en la ecuación (A3.2) obtenemos

$$J_n(jx) = \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{jx}{2}\right)^{n+2m}}{m!(n+m)!}$$

$$= j^n \sum_{m=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{n+2m}}{m!(n+m)!}$$

Advertimos ahora que $J_n(jx)$ multiplicada por una constante seguirá siendo una solución de la ecuación de Bessel. Por tanto, multiplicamos $J_n(jx)$ por la constante j^{-n}, obteniendo

$$j^{-n}J_n(jx) = \sum_{m=0}^{\infty} \frac{\left(\frac{1}{2}x\right)^{n+2m}}{m!(n+m)!}$$

Esta nueva función recibe el nombre de *función de Bessel modificada del primer tipo de orden n*, la cual se denota por medio de $I_n(x)$. En consecuencia, podemos expresar formalmente una solución de la ecuación de Bessel modificada, ecuación (A3.18), como

$$I_n(x) = j^{-n}J_n(jx)$$

$$= \sum_{m=0}^{\infty} \frac{\left(\frac{1}{2}x\right)^{n+2m}}{m!(n+m)!} \tag{A3.19}$$

La función de Bessel modificada $I_n(x)$ es una función real monotónicamente creciente del argumento x para toda n, como se indica en la figura A3.1 para $n = 0, 1, 2$.

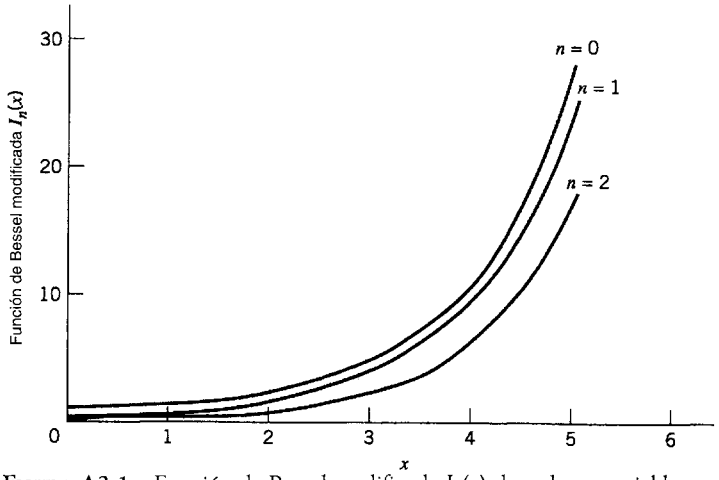


FIGURA A3.1 Función de Bessel modificada $I_n(x)$ de orden n variable.

La función de Bessel modificada $I_n(x)$ es idéntica a la función de Bessel original $J_n(x)$, salvo por una importante diferencia: los términos en la expansión en serie de la ecuación (A3.19) son en su totalidad positivos, en tanto que ellos alternan el signo en la expansión en serie de la ecuación (A3.2). La relación entre $J_n(x)$ e $I_n(x)$ es análoga a la forma en la cual las funciones trigonométricas cos x y sen x se relacionan con las funciones hiperbólicas cosh x y senh x.

Una propiedad interesante de la función de Bessel modificada $I_n(x)$ se deriva de la ecuación (A6.13). Específicamente, al sustituir x por jx y el ángulo ϕ por $\theta - \pi/2$, y recurriendo a la definición de $I_n(x)$ en la primera línea de la ecuación (A3.19), obtenemos

$$\sum_{n=-\infty}^{\infty} I_n(x)\ \exp(jn\theta) = \exp(x\ \cos\theta) \tag{A3.20}$$

A partir de esta relación se sigue que

$$I_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(x\ \cos\theta)\ \cos(n\theta)\ d\theta \tag{A3.21}$$

Esta fórmula integral para $In(x)$ puede, desde luego, obtenerse también de la ecuación (A3.4) realizando los cambios adecuados.

Cuando el argumento x es pequeño, obtenemos las siguientes estimaciones asintóticas directamente de la representación en serie de la ecuación (A3.19):

$$I_0(x) \to 1 \quad \text{para } x \to 0 \tag{A3.22}$$

e

$$I_n(x) \to 0 \quad \text{para } n \geq 1 \text{ y } x \to 0 \tag{A3.23}$$

Para grandes valores de x tenemos la siguiente estimación asintótica correspondiente a $I_n(x)$, la cual resulta válida para todos los enteros $n \geq 0$:

$$I_n(x) \simeq \frac{\exp(x)}{\sqrt{2\pi x}} \quad \text{para } x \to \infty \tag{A3.24}$$

Advierta que este comportamiento asintótico de $I_n(x)$ es independiente del orden n para grandes valores de x.

NOTAS Y BIBLIOGRAFÍA

1. La ecuación (A3.1) recibe su nombre en honor al matemático y astrónomo alemán Friedrich Wilhelm Bessel (1784-1846). En cuanto a tratamientos detallados de la solución de esta ecuación y de aspectos relacionados, vea Wylie y Barrett [174] y Watson [163].

APÉNDICE 4

FUNCIONES HIPERGEOMÉTRICAS CONFLUENTES

A4.1 *Ecuación de Kummer*

La *función hipergeométrica confluente*[1] es una solución de la *ecuación diferencial de Kummer:*

$$x\frac{d^2y}{dx^2}+(b-x)\frac{dy}{dx}-ay=0 \tag{A4.1}$$

donde, en general los parámetros a y b son números complejos. Para el caso en el que $b \neq 0, -1, -2, \ldots$, la solución de la ecuación de Kummer se define mediante la serie

$${}_1F_1(a;\ b;\ x)=1+\frac{a}{b}\frac{x}{1!}+\frac{a(a+1)}{b(b+1)}\frac{x^2}{2!}+\cdots \tag{A4.2}$$

donde ${}_1F_1(a; b; x)$ denota una función hipergeométrica confluente parametrizada por medio de a y b. En esta notación el primer subíndice denota el número de factoriales en el numerador del término general en la ecuación(A4.2), y el segundo subíndice indica el número de factoriales, además de $n!$, en el denominador. En la ecuación (A4.2) ambos subíndices son claramente 1.

A4.2 *Propiedades de la función hipergeométrica confluente*

Propiedad 1

Para valores pequeños de x, la función hipergeométrica confluente se aproxima como

$${}_1F_1(a;b;x)\simeq 1+\frac{a}{b}x \quad \text{para } x\to 0 \tag{A4.3}$$

Esta propiedad sigue directamente de la expansión en serie de la ecuación (A4.2).

Propiedad 2

Para $a = -1$ y $b = 1$ tenemos la identidad exacta:

$${}_1F_1(-1;\ 1;\ x)=1-x \quad \text{para toda } x \tag{A4.4}$$

Esta propiedad surge también directamente de la expansión en serie de la ecuación (A4.2).

Propiedad 3

La función hipergeométrica confluente para $a = -1/2$ y $b = 1$ *se relaciona exactamente con la función de Bessel modificada para toda* x *del modo siguiente:*

$$ {}_1F_1\left(-\frac{1}{2};\ 1;\ -x\right) = \exp\left(-\frac{x}{2}\right)\left((1+x)I_0\left(\frac{x}{2}\right) + xI_2\left(\frac{x}{2}\right)\right) \tag{A4.5} $$

donde $I_n(x)$ *es la función de Bessel modificada de orden* n.

Un caso especial de la ecuación (A4.5) ocurre cuando x es grande. A partir de la definición de la función de Bessel modificada dada en el apéndice 3, tenemos la siguiente fórmula asintótica para x grande:

$$ I_n(x) \simeq \frac{\exp(x)}{\sqrt{2\pi x}} \quad \text{para } x \to \infty \tag{A4.6} $$

De ese modo, al combinar las ecuaciones (A4.5) y (A4.6), obtenemos el resultado simple

$$ {}_1F_1\left(-\frac{1}{2};\ 1;\ -x\right) \simeq 2\sqrt{\frac{x}{\pi}} \quad \text{para } x \to \infty \tag{A4.7} $$

NOTAS Y BIBLIOGRAFÍA

1. En lo referente a una explicación de las funciones hipergeométricas confluentes, consulte Jeffreys y Jeffreys [74]. Los valores tabulados de estas funciones se presentan en Abramowitz y Stegun [1].

APÉNDICE 5

CRIPTOGRAFÍA

La privacidad es en verdad relevante para la seguridad e integridad de la transmisión de información. En realidad, la necesidad de contar con transmisiones seguras es más importante que nunca, en especial porque la dirección de gran parte de nuestros asuntos comerciales, de negocios y personales se está realizando en la actualidad por medio de las computadoras, mismas que han sustituido al medio tradicional del papel.

La *criptología* es el término que describe la ciencia de las comunicaciones secretas; se deriva del griego *kryptos* y *logos*, que significan, respectivamente, "oculto" y "palabra".[1] El tema de la criptología puede dividirse claramente en *criptografía* y *criptoanálisis*. La primera tiene que ver con las transformaciones de un mensaje en forma codificada mediante el *encriptado* y en la recuperación del mensaje original a partir del *desencriptado*. El mensaje original que se va a encriptar (cifrar) recibe el nombre de *texto común*, y el resultado que produce el encriptado se conoce como un *criptograma* o *texto cifrado;* los últimos dos términos se emplean indistintamente. El conjunto de transformaciones de datos para realizar el encriptado recibe el nombre de un *cifrador*; normalmente, las transformaciones se parametrizan mediante una o más *claves*. El *criptoanálisis,* por otra parte, tiene que ver con cómo deshacer las comunicaciones criptográficas rompiendo un cifrado o falseando señales codificadas que podrían ser aceptadas como genuinas.

Los sistemas criptográficos ofrecen tres servicios importantes:

1. *Privacía,* que se refiere a negar el acceso a la información a usuarios no autorizados.
2. *Autenticidad,* que se refiere a la validación de una fuente de mensaje.
3. *Integridad,* que se refiere a asegurar que el mensaje no sea modificado por medios accidentales o deliberados durante el tránsito.

Un sistema criptográfico convencional se sustenta en el uso de una pieza simple de información privada y necesariamente secreta conocida como la *clave*; por tanto, la criptografía convencional se conoce como *criptografía de clave simple o criptografía de clave secreta.*[2] Esta forma de criptografía opera sobre la premisa de que la clave la conocen el encriptador (emisor) y el desencriptador (receptor), pero no otros; lo que se supone es que, luego de que se encripta el mensaje, resulta imposible (probablemente) realizar el desencriptado sin el conocimiento de la clave.

La *criptografía de clave pública,*[3] denominada también *criptografía de dos claves,* difiere de la criptografía convencional con respecto a que ya no se trata de una sola clave secreta que comparten dos usuarios. Más bien, a cada usuario se le proporciona material de clave propio, y éste se divide en dos porciones: una componente pública y una privada. La pública genera una transformación pública, y la privada produce una transformación privada. Sin embargo, la transformación privada debe mantenerse secreta para mantener la comunicación entre los dos usuarios.

A5.1 *Criptografía de clave secreta*

Básicamente, el flujo de información en un sistema criptográfico de clave secreta es como se muestra en la figura A5.1. La fuente de mensaje genera un mensaje de texto común, que se encripta en un criptograma en el extremo de transmisión del sistema. El criptograma se envía a un *usuario autorizado* en el extremo de recepción por un canal "inseguro"; se considera que un canal es inseguro si su

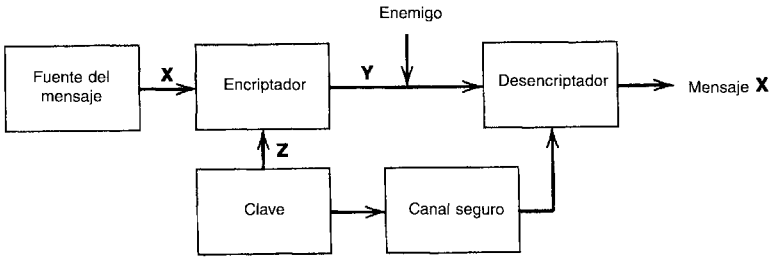


FIGURA A5.1 Diagrama de bloques del sistema criptográfico de clave secreta.

protección es inadecuada para las necesidades de sus usuarios. Se supone que en el curso de la transmisión, el criptograma puede ser interceptado por un *criptoanalista enemigo*[4] (es decir, un supuesto intruso en un sistema criptográfico). Lo que se requiere al respecto es efectuar el encriptado de manera tal que se evite que el enemigo aprenda los contenidos del mensaje de texto común.

En términos abstractos, un *sistema criptográfico o cifrado (para abreviar) se define como un conjunto de transformaciones invertibles del espacio del texto común (esto es, el conjunto de posibles mensajes de texto común) en el espacio del criptograma (es decir, el conjunto de todos los posibles criptogramas)*. Cada transformación particular corresponde a la encriptación (cifrado) de un texto común con una clave particular. La invertibilidad de la transformación significa que es posible un desencriptado único (descifrado) del criptograma cuando se conoce la clave. Sea **X** el mensaje de texto común, **Y** el criptograma y **Z** la clave. Considere que **F** denota la transformación invertible que produce el criptograma **Y**, como sigue:

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}, \mathbf{Z}) = \mathbf{F}_{\mathbf{Z}}(\mathbf{X}) \tag{A5.1}$$

La transformación busca hacer que el criptograma **Y** sea inútil para el enemigo. En el extremo receptor del sistema, el criptograma **Y** se desencripta con la transformación inversa $\mathbf{F}^{-1}$ para recuperar el mensaje de texto común **X**, de acuerdo con

$$\mathbf{F}^{-1}(\mathbf{Y}, \mathbf{Z}) = \mathbf{F}_{\mathbf{Z}}^{-1}(\mathbf{Y}) = \mathbf{F}_{\mathbf{Z}}^{-1}(\mathbf{F}_{\mathbf{Z}}(\mathbf{X})) = \mathbf{X} \tag{A5.2}$$

En términos físicos, el sistema criptográfico consiste en un conjunto de instrucciones, una pieza de hardware físico o un programa de computadora. En cualquier caso, el sistema se diseña para tener la capacidad de encriptar el texto común (y, desde luego, de desencriptar el criptograma resultante) de varias maneras. La forma particular elegida para realizar el encriptado real se determina mediante la clave específica.

La seguridad del sistema reside en la naturaleza secreta de la clave, lo cual requiere que ésta deba entregarse al receptor a través de un *canal seguro* (por ejemplo, correo certificado o servicio de mensajería), como se indica en la figura A5.1. El sistema criptográfico que se describe en esta figura proporciona una solución al *problema de privacía*, evitando que un enemigo extraiga información del mensaje transmitido por un canal de comunicación inseguro. La criptografía proporciona también una solución al *problema de autentificación* evitando que un criptoanalista enemigo finja ser el emisor del mensaje. En esta segunda situación, el criptoanalista enemigo es el que origina un "criptograma fraudulento" **Y′** que se entrega al receptor (del encriptador), como se muestra en la figura A5.2. El criptograma auténtico **Y** se muestra como una entrada punteada al criptoanalista enemigo, lo que indica que éste produce un criptograma fraudulento **Y′** sin haber visto el auténtico. El receptor puede ser capaz de reconocer **Y′** como fraudulento al desencriptarlo con la clave correcta **Z**; por tanto, la línea desde la salida del receptor hasta el destino se dibuja punteada para sugerir el rechazo del criptograma fraudulento **Y′** por el usuario receptor.

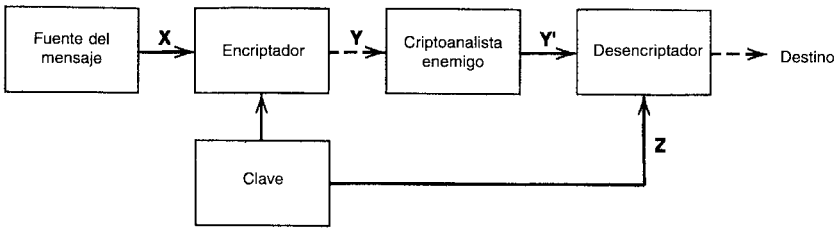


FIGURA A5.2 Ilustración de la intrusión de un criptoanalista enemigo.

A5.2 *Cifradores de bloque y de secuencia*

Muchos códigos de corrección de errores se clasifican en códigos de bloque y códigos convolucionales; los sistemas criptográficos (cifradores) pueden dividirse en dos amplias clases: *cifradores de bloque* y *cifradores de secuencia*. Los primeros operan de una manera puramente combinatoria sobre grandes bloques de texto común, en tanto que los de secuencia procesan el texto común en pedazos pequeños (es decir, caracteres o bits).

La figura A5.3 muestra la forma genérica de un cifrador de bloque. El texto común (consistente en datos en serie) se divide en grandes bloques, cada uno de los cuales suele integrarse con un número fijo de bits. Los bloques sucesivos de texto común se cifran (encriptan) utilizando la misma clave secreta, de manera independiente; los bloques cifrados que se producen se convierten finalmente en la forma serial. Así, un bloque de texto común particular idéntico a un bloque tal anterior origina un bloque de texto cifrado idéntico. De manera específica, cada bit de un bloque cifrado particular se elige como una función de todos los bits del bloque de texto común y la clave asociados; el objetivo de un cifrador de bloque es no tener un bit específico del texto común que aparezca alguna vez directamente en el texto cifrado.

Los cifradores de bloque operan con una transformación fija aplicada a grandes bloques de datos de texto común, en un esquema de bloque por bloque. En contraste, un cifrador de frecuencia opera con base en una transformación variable en el tiempo aplicada a bits individuales del texto común. Los cifradores de secuencia más populares son los denominados *cifradores de frecuencia aditivos binarios*, cuya forma genérica se muestra en la figura A5.4. En un cifrador de estas características, la clave secreta se usa para controlar un *generador de secuencia de* clave que emite una secuencia binaria llamada *de clave* cuya longitud es mucho mayor que la de la clave. Considere que x_n, y_n y z_n denotan el bit del texto común, el bit del texto cifrado y el bit de la secuencia de clave en el tiempo n, respectivamente. Los bits del texto cifrado se determinan entonces mediante la simple suma de módulo 2 de los bits de texto común y los bits de la secuencia de clave, como se indica por medio de

$$y_n = x_n \oplus z_n, \quad n = 1, 2, \ldots, N \tag{A5.3}$$

donde N es la longitud de la secuencia de clave. Debido a que la suma y la sustracción en la aritmética módulo 2 son exactamente iguales, la ecuación (A5.3) implica también la siguiente relación

$$x_n = y_n \oplus z_n, \quad n = 1, 2, \ldots, N \tag{A5.4}$$

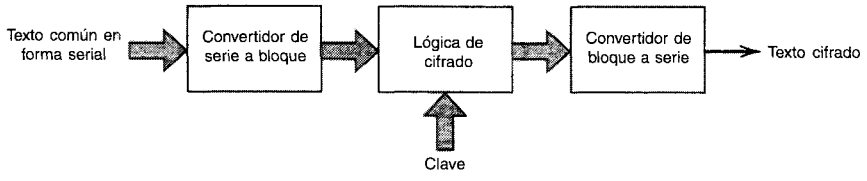


FIGURA A5.3 Diagrama de bloques de un cifrado de bloque.

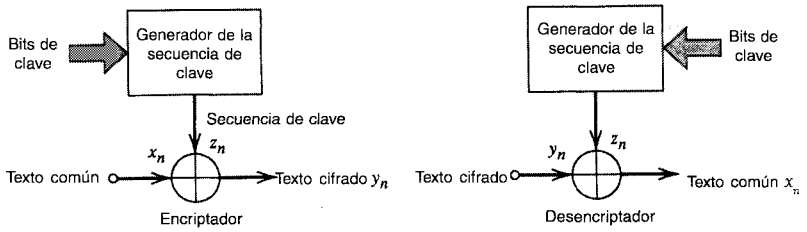


FIGURA A5.4 Cifrador de secuencia aditivo binario.

De este modo vemos que en cifradores de secuencia aditiva binarios es posible utilizar dispositivos idénticos para efectuar el encriptado y el desencriptado, como se muestra en la figura A5.4. La clave secreta se elige de acuerdo con alguna distribución de probabilidad. Para proporcionar encriptado seguro, la secuencia de clave debe asemejarse a una secuencia de lanzamientos de moneda (es decir, completamente aleatoria) en la forma más próxima posible.

Los cifradores de bloque se diseñan normalmente de tal manera que un pequeño cambio en un bloque de entrada de texto común produce un cambio mayor en la salida resultante. Esta propiedad de *propagación de errores* de los cifradores de bloque es valiosa en la autentificación en el sentido de que hace improbable que un criptoanalista enemigo modifique los datos encriptados, a menos de que disponga de conocimiento de la clave. Por otro lado, un cifrador de frecuencia aditiva binario *no* tiene propagación de errores. El desencriptado de un bit distorsionado en el texto cifrado afecta solamente al bit correspondiente de la salida resultante.

Los cifradores de frecuencia son por lo general más adecuados para la transmisión segura de datos por canales de comunicación propensos a errores; se usan en aplicaciones donde las altas velocidades de datos (como en el video seguro) son un requerimiento o cuando resulta esencial un retardo mínimo de la transmisión.[5]

■ REQUERIMIENTOS PARA LA PRIVACÍA

En la criptografía, una suposición fundamental es que un criptoanalista enemigo tiene conocimiento del mecanismo completo que se utiliza para realizar el encriptado, con excepción de la clave secreta. Es posible identificar las siguientes formas de ataque que puede intentar el criptoanalista enemigo, dependiendo de la disponibilidad de conocimiento adicional:

1. El *ataque únicamente del texto cifrado* es un ataque criptoanalítico en el cual el criptoanalista enemigo tiene acceso a parte o a todo el texto cifrado.
2. El *ataque de texto común conocido* es un ataque criptoanalítico en el que el criptoanalista tiene conocimiento de algunos pares de texto cifrado-texto común formados con la clave secreta real.
3. El *ataque de texto común escogido* es aquel en el cual el criptoanalista enemigo es capaz de enviar cualquier mensaje común escogido y de recibir en respuesta el texto cifrado correcto para la clave secreta real.
4. El *ataque de texto cifrado* es aquel en el cual el criptoanalista enemigo es capaz de elegir un texto cifrado arbitrario y encontrar el resultado correcto para su desencriptado.

El ataque únicamente del texto cifrado ocurre con frecuencia en la práctica. En esta forma de ataque, un criptoanalista enemigo utiliza sólo el conocimiento de la estructura estadística del lenguaje que se usa (por ejemplo, en inglés la letra *e* ocurre con una probabilidad de 13%, y la letra *q* siempre es seguida por una *u)* y el conocimiento de algunas palabras probables (por ejemplo, una carta probablemente empiece con "estimado Sr./Sra."). Un ataque de texto común conocido puede ocurrir en virtud de los formatos de computadora estándar que se utilizan en los lenguajes de programación y en la generación de datos. En cualquier caso, el ataque

únicamente del texto cifrado se considera como la amenaza más débil a la cual puede estar expuesto un sistema criptográfico, y cualquier sistema que sucumba a él es considerado en consecuencia totalmente inseguro. Por tanto, para que un sistema criptográfico ofrezca privacía, debe ser inmune como mínimo a los ataques únicamente del texto cifrado; idealmente, también debe ser inmune a los ataques de texto común conocido.

A5.3 *Enfoque de la teoría de la información*

En el *modelo de criptografía de Shannon*, nombrado de esa manera en reconocimiento al memorable artículo que este autor escribió en 1949 acerca del enfoque teórico de la información para los sistemas secretos, se supone que el criptoanalista enemigo tiene tiempo y potencia de cómputo ilimitados. Pero el enemigo se restringe presumiblemente al ataque de sólo texto cifrado. El criptoanálisis en el modelo de Shannon se define como el proceso para determinar la clave secreta, dado el criptograma (textos cifrados) y las probabilidades *a priori* de los diversos textos comunes y claves. Se considera que se *rompe* la privacía del sistema cuando el criptoanalista enemigo realiza con éxito el desencriptado, obteniendo una solución *única* para el criptograma.[6]

Sea $\mathbf{X} = (X_1, X_2, ..., X_N)$ un mensaje de texto común de N bits, y $\mathbf{Y} = (Y_1, Y_2, ..., Y_N)$ el correspondiente criptograma de N bits; esto es, tanto el texto común como el criptograma tienen el mismo número de bits. Se supone que la clave secreta **Z** utilizada para construir el criptograma se establece de acuerdo con alguna distribución de probabilidad. La incertidumbre en torno a **X** se expresa mediante la entropía $H(\mathbf{X})$, y la incertidumbre en relación con **X** dado el conocimiento de **Y** se expresa mediante la entropía condicional $H(\mathbf{X}|\mathbf{Y})$. La *información mutua* entre **X** y **Y** se define mediante

$$I(\mathbf{X};\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \tag{A5.5}$$

La información mutua $I(\mathbf{X}; \mathbf{Y})$ representa una medida de seguridad (privacía) básica en el modelo de Shannon.

■ SEGURIDAD PERFECTA

Suponiendo que un criptoanalista enemigo puede observar sólo el criptograma **Y**, parece apropiado que definamos la *seguridad perfecta* de un sistema criptográfico para significar que el texto común **X** y el criptograma **Y** son estadísticamente independientes. En otras palabras, tenemos

$$I(\mathbf{X};\mathbf{Y}) = 0 \tag{A5.6}$$

En ese caso, utilizando la ecuación (A5.5), encontramos que la condición para la seguridad perfecta puede escribirse como

$$H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) \tag{A5.7}$$

La ecuación (A5.7) establece que lo máximo que puede hacer un criptoanalista enemigo, dado el criptograma **Y**, es suponer el mensaje del texto común **X** de acuerdo con la distribución de probabilidad de todos los mensajes posibles.

Dada la clave secreta **Z**, admitimos que

$$\begin{aligned} H(\mathbf{X}\,|\,\mathbf{Y}) &\le H(\mathbf{X}, \mathbf{Z}\,|\,\mathbf{Y}) \\ &= H(\mathbf{Z}\,|\,\mathbf{Y}) + H(\mathbf{X}\,|\,\mathbf{Y}, \mathbf{Z}) \end{aligned} \tag{A5.8}$$

La entropía condicional $H(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$ es cero si, y solamente si, **Y** y **Z** juntas determinan en forma única a **X**; ésta es realmente una suposición válida cuando el proceso de desencriptado se lleva a cabo con conocimiento de la clave secreta **Z**. Por tanto, es posible simplificar la ecuación (A5.8) como sigue:

$$\begin{aligned} H(\mathbf{X}|\mathbf{Y}) &\le H(\mathbf{Z}|\mathbf{Y}) \\ &\le H(\mathbf{Z}) \end{aligned} \tag{A5.9}$$

Por consiguiente, al sustituir la ecuación (A5.9) en la (A5.7) encontramos que para que un sistema criptográfico proporcione seguridad perfecta, debe satisfacerse la siguiente condición

$$H(\mathbf{Z}) \geq H(\mathbf{X}) \tag{A5.10}$$

La desigualdad de la ecuación (A5.10) es la *cota fundamental de Shannon para la seguridad perfecta*; ésta establece que para seguridad perfecta, la incertidumbre de la clave secreta **Z** debe ser al menos tan grande como la incertidumbre del texto común **X** que es encubierto por la clave.

En el caso de que los alfabetos del texto común y de la clave sean del mismo tamaño, el uso de la cota de Shannon para la seguridad perfecta produce el siguiente resultado: *la clave debe ser al menos tan larga como el texto común.* La conclusión que se saca de este resultado es que la longitud de la *clave secreta* que se necesita para conformar un sistema criptográfico perfectamente seguro quizá sea imprácticamente grande en la mayor parte de las aplicaciones. A pesar de eso, la seguridad perfecta tiene un lugar en el escenario práctico: puede usarse cuando el número de mensajes posibles sea pequeño o en casos en los que la mayor importancia se vincula con la seguridad perfecta.

Un cifrado perfectamente seguro bien conocido es el *bloc de un tiempo*[7] (algunas veces llamado el *cifrado Vernam)*, el cual se usa en aplicaciones no convencionales, como la de dos usuarios que se comunican en una línea activa con requerimientos de alta confidencialidad. El bloc de un tiempo es un cifrado de secuencia para el cual la clave es la misma que la secuencia de clave, como se indica en la figura A5.5. En el encriptado la entrada consiste en dos componentes: un mensaje representado por una secuencia de bits de mensaje $\{x_n | n = 1, 2, ...\}$, y una clave representada por una secuencia de bits independientes y distribuidos uniformemente $\{z_n | n = 1, 2, ...\}$. El cifrado resultante $\{y_n | n = 1, 2, ...\}$ se obtiene mediante la suma módulo 2 de las dos secuencias de entrada, como indica

$$y_n = x_n \oplus z_n, \quad n = 1, 2, \ldots$$

Considere, por ejemplo, la secuencia de mensaje binaria 00011010 y la secuencia de clave binaria 01101001. La suma módulo 2 de esta secuencia se escribe de la manera siguiente:

Mensaje: 00011010
Clave: 01101001
Cifrado: 01110011

En la regla de encriptado que se describe aquí, el bit de clave 1 intercambia ceros y unos en la secuencia de mensaje, y el bit de clave 0 deja los bits de mensaje inalterados. La secuencia de mensaje se recupera simplemente mediante la suma módulo 2 del cifrado binario y las secuencias de clave, como indica

Cifrado: 01110011
Clave: 01101001
Mensaje: 00011010

El bloc de un tiempo es perfectamente seguro, debido a que la información mutua entre el mensaje y el cifrador es cero; por tanto, es completamente indescifrable.

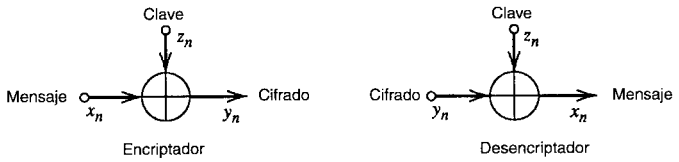


FIGURA A5.5 Bloque de un tiempo (cifrado de Vernam).

■ DISTANCIA DE UNICIDAD

Considere ahora el caso práctico de un cifrado imperfecto y haga la pregunta: ¿cuándo puede un criptoanalista enemigo romper el cifrado? A medida que aumenta la cantidad de texto interceptado, esperamos intuitivamente que sea posible alcanzar un punto en el cual se vuelva factible para un criptoanalista enemigo con tiempo y capacidad de cómputo ilimitados descubrir la clave y romper de este modo el cifrado. Este punto crítico en el modelo de Shannon recibe el nombre de la *distancia de unicidad,* la cual se define formalmente como la N más pequeña tal que la entropía condicional $H(\mathbf{Z}|Y_1, Y_2, ..., Y_N)$ sea aproximadamente cero. Para un tipo particular de "cifrado aleatorio", la distancia de unicidad está dada aproximadamente por[8]

$$N_0 \simeq \frac{H(\mathbf{Z})}{r \log L_y} \tag{A5.11}$$

donde $H(\mathbf{Z})$ es la entropía de la clave $\mathbf{Z}$, y L_y es el tamaño del texto cifrado. El parámetro r es la *redundancia porcentual* de la información del mensaje contenida en el texto cifrado de N bits, el cual a su vez está definido por

$$r = 1 - \frac{H(\mathbf{X})}{N \log L_y} \tag{A5.12}$$

donde $H(\mathbf{X})$ es la entropía del texto común $\mathbf{X}$. En casi todos los sistemas criptográficos, el tamaño L_y del alfabeto del texto cifrado es el mismo que el tamaño L_x del alfabeto del texto común; en tal caso, r es exactamente la redundancia porcentual del propio texto común. Aunque la derivación de la ecuación (A5.11) supone cierto "cifrador aleatorio" bien definido, ésta puede utilizarse para estimar la distancia de unicidad para tipos ordinarios de cifradores, lo cual constituye la práctica rutinaria actual.

Sea K el número de dígitos en la clave $\mathbf{Z}$ que se elige de un alfabeto de tamaño L_z; entonces podemos expresar la entropía de la clave $\mathbf{Z}$ en la forma siguiente:

$$H(\mathbf{Z}) \leq \log(L_z^K) = K \log L_z \tag{A5.13}$$

con la igualdad si y sólo si la clave es completamente aleatoria. Sea el tamaño L_z del alfabeto de clave el mismo que el tamaño L_y del alfabeto del texto cifrado, y elíjase la clave completamente al azar para maximizar la distancia de unicidad. En ese caso, al sustituir la ecuación (A5.13) con la igualdad en la ecuación (A5.11), obtenemos el resultado simple

$$N_0 \simeq \frac{K}{r} \tag{A5.14}$$

Para ilustrar la aplicación de la ecuación (A5.14), considere un sistema criptográfico con $L_x = L_y = L_z$, el cual se usa para el encriptado de texto en inglés. La redundancia porcentual r para el texto inglés característico es casi de 75%. Por tanto, de acuerdo con la ecuación (A5.14), un criptoanalista enemigo puede romper el cifrado después de interceptar únicamente cerca de 1.333 K bits de texto cifrado, donde $\mathbf{K}$ es el tamaño de la clave.

Sin embargo, es importante advertir que un cifrado imperfecto que potencialmente puede romperse es, en algunos casos, de valor práctico. Cuando el texto cifrado que se intercepta contiene información suficiente para satisfacer la ecuación (A5.11), no existe garantía de que un criptoanalista enemigo con recursos de cómputo limitados tenga la posibilidad de romper realmente el cifrado. De modo específico, es posible que el cifrador sea diseñado de tal manera que la tarea del criptoanálisis, si bien se sabe puede lograrse con una cantidad finita de cómputo, resulta tan abrumadora que literalmente acaba con los recursos de cómputo físicos del universo. En ese caso, se afirma que el cifrado imperfecto será *seguro computacionalmente.*

Papel de la compresión de datos en la criptografía

La compresión de datos sin pérdida o la compactación de datos constituye una herramienta útil en la criptografía. Afirmamos esto debido a que la compactación de datos elimina la redundancia, incrementando de ese modo la distancia de unicidad N_0 de acuerdo con la ecuación (A5.11). Para explotar esta idea se usa la compactación de datos antes del encriptado en el transmisor, y la información redundante se reinserta después del desencriptado en el receptor. El resultado neto es que el usuario autorizado a la salida del receptor no ve diferencia, e incluso se ha hecho más segura la transmisión de información. Sería tentador considerar el uso de la compactación de datos perfecta para eliminar toda la redundancia, transformando de esa forma una fuente de mensaje en una fuente completamente aleatoria y dando lugar a $N_0 = \infty$ con cualquier tamaño de clave. Desafortunadamente, no contamos con un dispositivo capaz de efectuar la compactación de datos perfecta en fuentes de mensajes realistas, ni es probable que alguna vez exista un dispositivo de este tipo. Por tanto, resulta vano confiar exclusivamente en la compactación de datos para la seguridad de los mismos. A pesar de eso, la compactación de datos limitada tiende a aumentar la seguridad, lo cual es la razón por la que los criptógrafos consideran la compresión de datos como una estratagema útil.

Difusión y confusión

En el modelo de Shannon de la criptografía se sugieren dos métodos por sí solos como principios generales para guiar el diseño de cifradores prácticos. Éstos reciben el nombre de *difusión* y *confusión*, cuyos objetivos (propio o conjunto) consisten en frustrar un análisis estadístico del texto cifrado por parte del enemigo y, por tanto, en hacer extremadamente difícil la ruptura del cifrado.

En el método de difusión, la estructura estadística del texto común se oculta al dispersar la influencia de un solo bit en el texto común sobre un gran número de bits en el texto cifrado. Esta dispersión tiene el efecto de forzar al enemigo a interceptar una cantidad tremenda de material para la determinación de la estructura estadística del texto común, ya que la estructura es únicamente evidente en muchos bloques, cada uno de los cuales tiene una probabilidad muy pequeña de ocurrencia. En el método de confusión, las transformaciones de datos se diseñan para complicar la determinación de la forma en la cual las estadísticas del texto cifrado dependen de las estadísticas del texto común. Así, un buen cifrado usa una combinación de la difusión y la confusión.

Sin embargo, para que un cifrado sea de valor práctico no sólo debe ser difícil que el criptoanalista enemigo lo rompa, sino que también debe ser fácil encriptar y desencriptar los datos dado el conocimiento de la clave secreta. Podemos satisfacer estos dos objetivos de diseño utilizando un *cifrado de producto* basado en la idea de "divide y vencerás". Específicamente, la puesta en práctica de un cifrado sólido se logra como una sucesión de cifrados componentes simples, cada uno de los cuales contribuye con una cantidad modesta de difusión y confusión para conformar el cifrado completo. Los cifrados de producto suelen integrarse utilizando como componentes básicos los cifrados de sustitución y los de transposición; estos últimos se describen a continuación.

1. *Cifrado de sustitución.*
En un cifrado de este tipo cada letra del texto común se reemplaza por un sustituto fijo, usualmente también una letra del mismo alfabeto, con la regla de sustitución particular determinándose mediante la clave secreta. De ese modo el texto común

$$\mathbf{X} = (x_1, x_2, x_3, x_4, \ldots)$$

donde x_1, x_2, x_3, ..., son las letras sucesivas, se transforma en el texto cifrado

$$\begin{aligned} \mathbf{Y} &= (y_1, y_2, y_3, y_4, \ldots) \\ &= (f(x_1), f(x_2), f(x_3), f(x_4), \ldots) \end{aligned} \tag{A5.15}$$

Letras de texto común	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Letras de texto cifrado	YDUBHNACSVXELPFMKQJRWGOZIT

FIGURA A5.6 Cifrado de sustitución.

donde $f(\cdot)$ es una función con una inversa. Cuando los sustitutos son letras, la clave es una permutación del alfabeto. Considere, por ejemplo, el alfabeto de texto cifrado de la figura A5.6, donde vemos que la primera letra *Y* es el sustituto de *A*, la segunda letra es el sustituto de *D*, etc. El uso de un cifrado de sustitución produce confusión.

2. *Cifrado de transposición.*
En un cifrado de transposición, el texto común se divide en grupos de periodo fijo d y la misma permutación se aplica a cada grupo, con la regla de permutación particular determinándose mediante la clave secreta. Por ejemplo, considere la regla de permutación descrita en la figura A5.7, para la cual el periodo es $d = 4$. De acuerdo con este cifrado, la letra x_1 se mueve de la posición 1 del texto común a la posición 4 del texto cifrado. De ese modo, el texto común

$$\mathbf{X} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, \ldots)$$

se transforma en el texto cifrado

$$\mathbf{Y} = (x_3, x_4, x_2, x_1, x_7, x_8, x_6, x_5, \ldots)$$

Aunque las estadísticas de una sola letra de texto cifrado **Y** son las mismas que las de texto común **X**, se alteran las estadísticas de orden superior. El uso de un cifrado de transposición origina difusión.

Al intercalar las sustituciones y transposiciones simples y repetir el proceso de intercalamiento muchas veces, es posible conformar un sólido cifrado que cuente con difusión y confusión adecuadas.

▶ EJEMPLO A5.1

Considere el mensaje de texto común

THE KING IS DEAD LONG LIVE THE KING

Utilizando el alfabeto permutado que se describe en la figura A5.6 para el cifrado de sustitución, este texto común se transforma en el texto cifrado

RCHXSPASJBHYBEFPAESGHRCHXSPA

Suponga ahora que aplicamos la regla de permutación que se describe en la figura A5.7 para el cifrado de transposición; por tanto, el texto cifrado que resulta del cifrado de sustitución se transforma adicionalmente en

HXCRASPSHYBJFBEBSGEACHRHPASX

el cual no se asemeja al texto común original. ◀

Letras de texto común	x_1	x_2	x_3	x_4
Letras de texto cifrado	x_3	x_4	x_2	x_1

FIGURA A5.7 Cifrado de transposición.

A5.4 *Estándar de encriptado de datos*

El *estándar de encriptado de datos* (DES)[9] es ciertamente el criptoalgoritmo de clave secreta mejor conocido y discutiblemente el más usado; el término *algoritmo* se utiliza para describir una secuencia de cálculos. El algoritmo DES básico puede utilizarse tanto para el encriptado de datos como para la autentificación de los mismos. Se trata de un criptoalgoritmo estándar para los sistemas de almacenamiento de datos y correo, transferencias electrónicas de fondos (menudeo y mayoreo) y el intercambio electrónico de datos de negocios.

El algoritmo DES es un sólido cifrado de bloques que opera en bloques de 64 bits de datos de texto común y utiliza una clave de 56 bits; se diseña de acuerdo con los métodos de difusión y confusión de Shannon. Esencialmente se utiliza el mismo algoritmo para el encriptado y el desencriptado. Las transformaciones completas que se emplean en el algoritmo DES pueden escribirse como $P^{-1}\{F[P(\mathbf{X})]\}$, donde $\mathbf{X}$ es el texto común, P es cierta permutación y la función F combina constituciones y transposiciones. La función F se obtiene a su vez al poner en cascada cierta función f, con cada etapa de la cascada referida como una *ronda*.

El diagrama de flujo de la figura A5.8 muestra los detalles del algoritmo DES para el encriptado. Después de cierta permutación inicial, un texto común de 64 bits se divide en la mitad izquierda L_0 y en la mitad derecha R_0, cada una de las cuales con 32 bits de largo. El algoritmo lleva a cabo entonces 16 rondas de cálculo dependiendo de la clave, con la i-ésima ronda del cálculo descrita de la manera siguiente:

$$L_i = R_{i-1} \qquad i = 1, 2, \ldots, 16 \tag{A5.16}$$

$$R_i = L_{i-1} \oplus f(R_{i-1}, Z_i) \qquad i = 1, 2, \ldots, 16 \tag{A5.17}$$

En el segundo miembro de la ecuación (A5.17), la suma es módulo 2 y cada Z_i es un bloque diferente de 48 bits de la clave utilizada en la ronda i. La función $f(\cdot, \cdot)$ tiene salida de 32 bits. El resultado de la 16ava ronda se invierte, obteniendo la secuencia $R_{16}L_{16}$. Esta secuencia de 32 bits se introduce en una permutación final P^{-1} para producir el texto cifrado de 64 bits. El objetivo es que después de 16 rondas de cálculos dependientes de la clave, los patrones en el texto común original resultan indetectables en el texto cifrado. A partir de las ecuaciones (A5.16) y (A5.17), notamos que para el desencriptado, la función $f(\cdot, \cdot)$ no necesita ser invertible, ya que (L_{i-1}, R_{i-1}) puede recuperarse de (L_i, R_i) simplemente del modo siguiente:

$$R_{i-1} = L_i \qquad i = 1, 2, \ldots, 16 \tag{A5.18}$$

$$L_{i-1} = R_i \oplus f(L_i, Z_i) \qquad i = 1, 2, \ldots, 16 \tag{A5.19}$$

La ecuación (A5.19) se cumple incluso si la función $f(\cdot, \cdot)$ es una función de varios a uno (es decir, no tiene una inversa única).

La figura A5.9 presenta el diagrama de flujo correspondiente al cálculo de la función $f(\cdot, \cdot)$. El bloque R de 32 bits se expande primero en un nuevo bloque R' de 48 bits repitiendo los bits de los extremos de cada palabra sucesiva de 4 bits (esto es, los bits numerados 1, 4, 5, 8, 9, 12, 13, 16, ..., 28, 29, 32). Por tanto, dado el bloque R de 32 bits escrito como

$$R = \underbrace{r_1 r_2 r_3 r_4}_{\substack{\text{primera palabra} \\ \text{de 4 bits}}} \quad \underbrace{r_5 r_6 r_7 r_8}_{\substack{\text{segunda palabra} \\ \text{de 5 bits}}} \quad \cdots \quad \underbrace{r_{29} r_{30} r_{31} r_{32}}_{\substack{\text{octava palabra} \\ \text{de 4 bits}}}$$

construimos el bloque R' expandido de 48 bits en la forma:

$$R' = \underbrace{r_{32} r_1 r_2 r_3 r_4 r_5}_{\substack{\text{primera palabra} \\ \text{de 6 bits}}} \quad \underbrace{r_4 r_5 r_6 r_7 r_9}_{\substack{\text{segunda palabra} \\ \text{de 6 bits}}} \quad \cdots \quad \underbrace{r_{28} r_{29} r_{30} r_{31} r_{32} r_1}_{\substack{\text{octava palabra} \\ \text{de 6 bits}}}$$

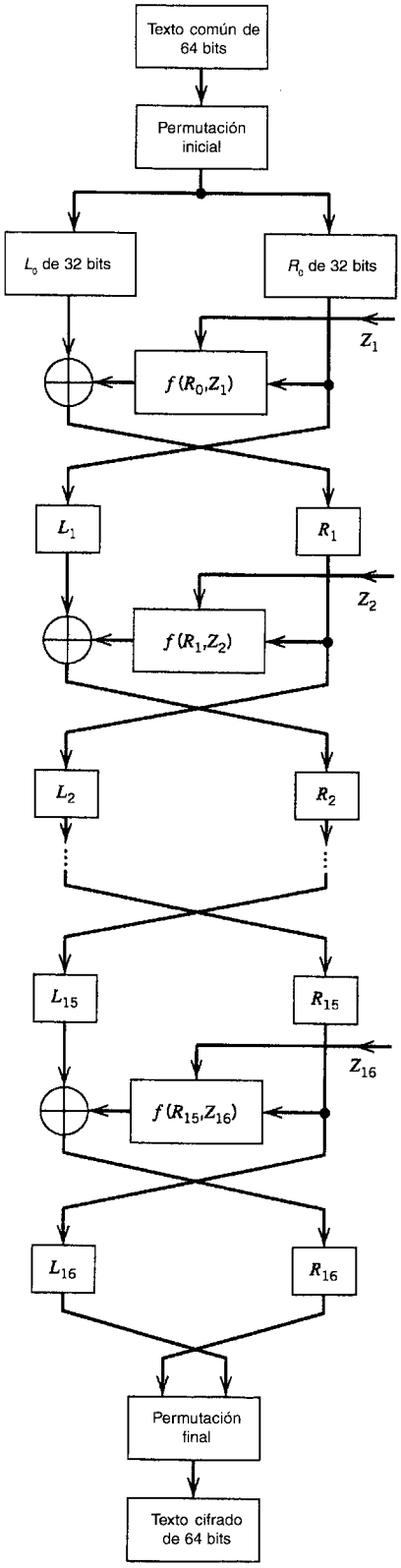


FIGURA A5.8 Estándar de encriptado de datos. (Tomado de Diffie y Hellman [234] con autorización del IEEE.)

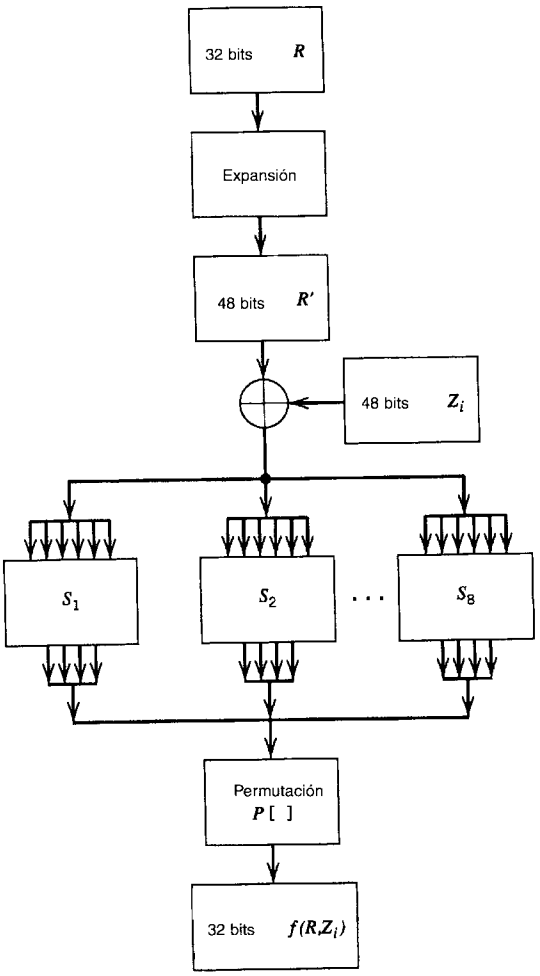


FIGURA A5.9 Diagrama de flujo $f(R, K)$. (Tomado de Diffie y Hellman [234] con autorización del IEEE.)

Los bloques R' y Z_i de 48 bits se someten a una suma módulo 2 y la resultante se divide en ocho palabras de 6 bits. Considere que estas palabras se denotan mediante B_1, B_2, ..., B_8. De ese modo escribimos

$$B_1B_2 \ldots B_8 = R' \quad \oplus Z_i \tag{A5.20}$$

Cada palabra B_i de 6 bits se alimenta a una caja de sustitución S_i en la forma de una tabla de búsqueda, produciendo una salida $S_i(B_i)$ de 4 bits. Cada bit de salida de la caja de sustitución $S_i(B_i)$ es una función booleana de la palabra B_i de 6 bits. Las ocho salidas $S_1(B_1)$, $S_2(B_2)$, ..., $S_8(B_8)$ se arreglan en un bloque sencillo de 32 bits que es la entrada para la caja de permutación denotada por $P[\cdot]$. La salida permutada que se produce de esta manera es la función deseada de 32 bits $f(R, Z_i)$, como se indica mediante

$$f(R, Z_i) = P[S_1(B_1)S_2(B_2) \ldots S_8(B_8)] \tag{A5.21}$$

El bloque Z_i de 48 bits para la iteración i-ésima utiliza un conjunto diferente de la clave de 64 bits Z_0. El procedimiento que se utiliza para determinar cada Z_i se conoce como *cálculo de la lista de claves*, cuyo diagrama de flujo se presenta en la figura A5.10. La clave Z_0 tiene 8 bits de paridad en las posiciones 8, 16, ..., 64, los cuales se usan para la detección de errores en sus bytes

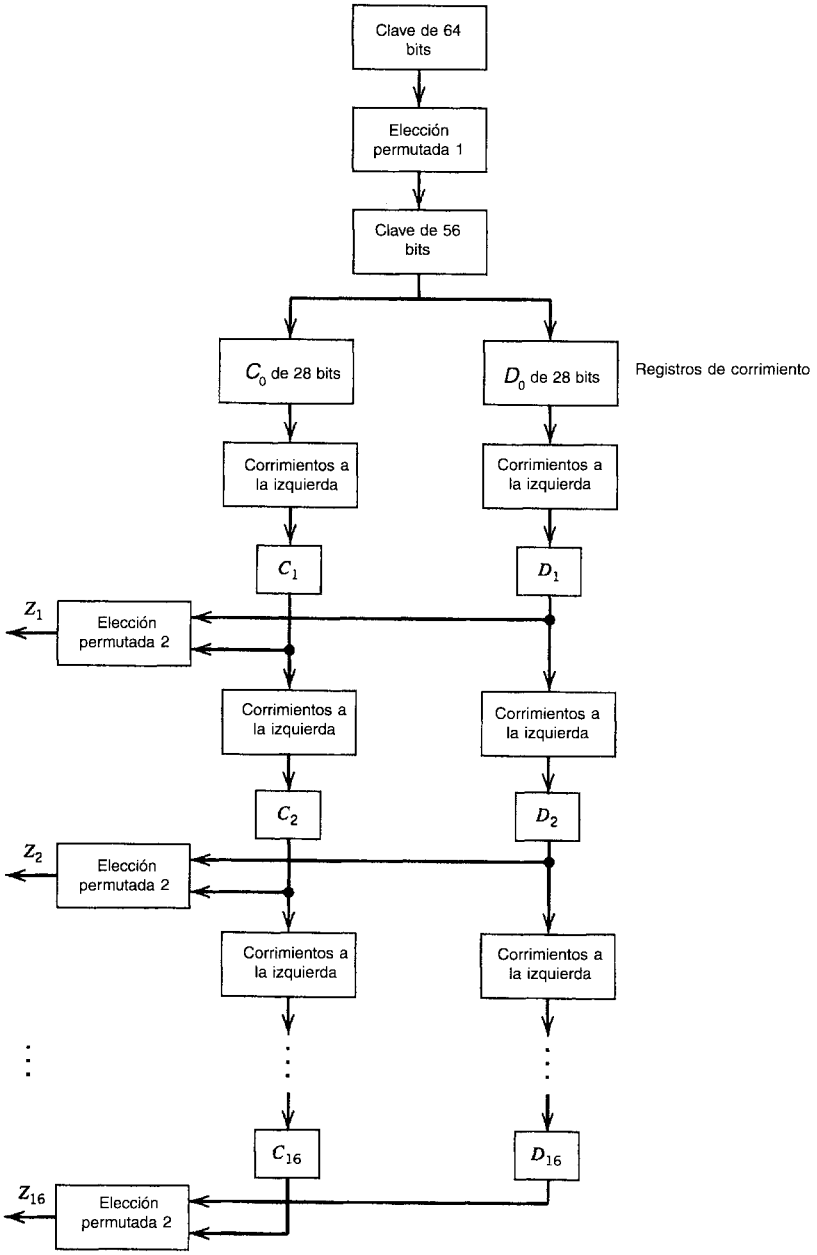


FIGURA A5.10 Diagrama de flujo para el cálculo de la lista de claves. (Tomado de Diffie y Hellman [234] con autorización del IEEE.)

respectivos de 8 bits; los errores de interés pueden surgir en la generación, distribución y almacenamiento de la clave Z_0. La elección permutada 1 descarta los bits de paridad Z_0 y luego permuta los 56 bits restantes que se cargan en dos registros de corrimiento de 28 bits, cada uno con 24 derivaciones. Las 48 derivaciones de los dos registros de corrimiento se someten a 16 iteraciones de cálculo, incluyendo cada una de ellas 1 o 2 corrimientos izquierdos cíclicos seguidos por una permutación, la cual se denomina la elección 2.

Las salidas que resultan de estas 16 iteraciones proporcionan los diferentes bloques de 48 bits $Z_1, Z_2, \ldots, Z_{16}$ de la clave utilizada en la iteración 1, 2, ..., 16, respectivamente.

A pesar de todas las demandas en contra, es manifiesto que ninguna ha demostrado aún una debilidad fundamental del algoritmo DES. A pesar de toda la controversia en torno a su uso, quizá la aportación más importante del algoritmo DES es el hecho de que ha sido útil para aumentar el nivel de interés en el uso de la criptografía como un mecanismo para las redes de cómputo seguras.

A5.5 *Criptografía de clave pública*[10]

Para que un par de usuarios participen en la comunicación criptográfica por un canal inseguro, es necesario que éstos intercambien información de clave antes de la comunicación. El requerimiento para una distribución segura de claves entre usuarios autorizados se aplica en todos los sistemas criptográficos, sin que importe su tipo. En la criptografía convencional, los usuarios emplean un canal físicamente seguro (por ejemplo, servicios de mensajería o correo certificado) para la distribución de claves. Sin embargo, el uso de un canal suplementario de este tipo señala una limitación mayor de la criptografía convencional. Como se sabe, el uso de servicio de mensajería o correo certificado para la distribución de claves es costoso, inconveniente, de ancho de banda bajo y lento; además no siempre es seguro.

El problema de la distribución de claves se acentúa en particular en las grandes redes de comunicación, donde el número de posibles conexiones crece como $(n^2-n)/2$ para n usuarios. Para n grande, el costo de la distribución de claves se vuelve prohibitivo. Por tanto, en el desarrollo de redes de comunicaciones grandes y seguras, estamos obligados a confiar en el uso de canales inseguros tanto para el intercambio de información de clave como para la subsecuente comunicación segura. Esta restricción origina una pregunta fundamental: ¿cómo puede intercambiarse en forma segura la información de clave por un canal inseguro? En la *criptografía de clave pública* este aspecto aparentemente difícil se resuelve haciendo que una parte del material de clave sea "público", lo cual simplifica considerablemente la tarea de la administración de claves. Esto contrasta directamente con la criptografía convencional, donde la clave se mantiene completamente oculta para un criptoanalista enemigo.

Un sistema criptográfico de clave pública está descrito por dos conjuntos de algoritmos que calculan las funciones invertibles (transformaciones). Dichos conjuntos se denotan por medio de $\{E_z\}$ y $\{D_z\}$ que se indexan mediante z. Las transformaciones calculadas por estos algoritmos pueden escribirse como sigue:

$$E_z : f_z(x) = y \tag{A5.22}$$

$$D_z : f_z^{-1}(y) = x \tag{A5.23}$$

donde x es cierto mensaje de entrada en el dominio de alguna función f_z indexada por z, y y es el criptograma correspondiente en el intervalo de f_z. Un requerimiento fundamental del sistema es que f_z debe ser una *función de trampa de una vía*. El término "una vía" se refiere al hecho de que para x en el dominio de f_z, debe ser fácil calcular $f_z(x)$ a partir del conocimiento del algoritmo E_z, pero para cierto criptograma y en el intervalo de f_z, un criptoanalista enemigo debe encontrar sumamente difícil calcular la inversa $f_z^{-1}(y)$. Por otra parte, un usuario autorizado que tenga posesión del algoritmo asociado D_z encontraría fácil calcular la inversa $f_z^{-1}(y)$. De esa manera, la *clave privada* (algoritmo) D_z ofrece una "trampa" que provoca que el problema f_z parezca extremadamente difícil desde las perspectivas del criptoanalista, aunque fácil para el poseedor (único autorizado) de D_z. Puesto que el conocimiento de la clave (algoritmo) E_z por sí solo no hace posible el cálculo de la inversa f_z, éste puede hacerse *público*; de ahí el nombre "criptografía de clave pública".

La noción que surge de la descripción de un sistema criptográfico de clave pública presentado aquí es que las claves vienen en pares inversos (es decir, clave pública y clave privada) y que cada par de claves tiene dos propiedades básicas:

1. *Cualquier mensaje que se encripte con una de las claves puede desencriptarse con la otra clave.*
2. *Dado el conocimiento de la clave pública, no es computacionalmente factible encontrar la clave secreta.*

El uso de la criptografía de clave pública según se describe aquí posibilita resolver el problema de privacía de la forma siguiente. Los suscriptores a un sistema de comunicación seguro listan sus claves públicas en un "directorio telefónico" junto con sus nombres y direcciones. Un suscriptor puede enviar entonces un mensaje privado a otro suscriptor buscando simplemente la clave pública del destinatario y utilizando la clave para encriptar el mensaje. El mensaje encriptado (es decir, el texto cifrado), sólo puede leerlo el poseedor de esa clave en particular. De hecho, si se perdiera el mensaje original (esto es, el texto común), incluso su emisor encontraría en extremo difícil recuperarlo del texto cifrado.

La administración de claves de la criptografía de clave pública permite que ésta resulte apropiada para el desarrollo de grandes redes de comunicación seguras. En realidad, ésta ha evolucionado desde un concepto simple hasta un soporte principal de la tecnología criptográfica.

■ DISTRIBUCIÓN DE CLAVE PÚBLICA DE DIFFIE-HELLMAN

En un sistema simple e incluso elegante conocido como *sistema de distribución de clave pública de Diffie-Hellman* se aprovecha el hecho de que es fácil calcular una exponencial discreta, pero difícil calcular un logaritmo discreto. Para ser más específicos, considere la *función exponencial discreta*

$$Y = \alpha^X \bmod p \qquad \text{para } 1 \leq X \leq p-1 \tag{A5.24}$$

donde la aritmética que se realiza es módulo p. La α es un entero que sería *primitivo* (es decir, todas las potencias de alfa generan todos los elementos mod p relativamente primos a $p-1$). De manera correspondiente, X se conoce como el *logaritmo discreto* de Y para la base α, mod p, según indica

$$X = \log_\alpha Y \bmod p \qquad \text{para } 1 \leq Y \leq p-1 \tag{A5.25}$$

El cálculo de Y a partir de X es sencillo, utilizando el artificio de elevar al cuadrado y multiplicar. Por ejemplo, para $X = 16$ tenemos

$$Y = \alpha^{16} = \left\{[(\alpha^2)^2]^2\right\}^2$$

Por otra parte, el problema de calcular X a partir de Y es mucho más difícil.

En el sistema de distribución de clave pública de Diffie-Hellman, se supone que todos los usuarios conocen tanto α como p. Un usuario i, digamos, elige un número aleatorio independiente X_i uniformemente de un conjunto de enteros $\{1, 2, ..., p)$ que se mantiene como un *secreto privado*. Pero la exponencial discreta

$$Y_i = \alpha^{X_i} \bmod p \tag{A5.26}$$

se deposita en un *directorio público* con el nombre y dirección del usuario. Todos los demás usuarios del sistema hacen lo mismo. Ahora bien, suponga que los usuarios i y j desean comuni-

carse en forma privada. Para proceder, el usuario i va a buscar a Y_j en el directorio público y utiliza el secreto privado X_i para calcular

$$\begin{aligned} K_{ji} &= (Y_j)^{X_i} \bmod p \\ &= (\alpha^{X_j})^{X_i} \bmod p \\ &= \alpha^{X_j X_i} \bmod p \end{aligned} \tag{A5.27}$$

De manera similar, el usuario j calcula K_{ij}. Sin embargo, tenemos

$$K_{ji} = K_{ij} \tag{A5.28}$$

Por tanto, los usuarios i y j llegan a K_{ji} como la *clave secreta* en un criptosistema convencional. Otro usuario debe calcular K_{ij} utilizando la información Y_i y Y_j que se obtiene del directorio público, aplicando la fórmula alternativa

$$K_{ji} = (Y_j)^{\log Y_i} \bmod p \tag{A5.29}$$

Aparentemente, no existe otro método para que un enemigo descubra la clave secreta K_{ij}; sin embargo, no hay prueba de ello. De acuerdo con lo dicho antes, el cálculo de la ecuación (A5.29) resulta difícil, ya que incluye un logaritmo discreto, en tanto que la ecuación (A5.27) se calcula fácilmente, pues incluye una exponencial discreta. De ese modo, la seguridad del sistema depende de la dificultad que se encuentra en el cálculo de un logaritmo discreto.

El sistema de distribución de clave pública de Diffie-Hellman es el más antiguo de su clase; no obstante, sigue considerándosele generalmente como uno de los sistemas de distribución de clave pública más seguros y prácticos.

A5.6 *Sistema Rivest-Shamir-Adleman*

Desarrollar un sistema criptográfico de clave pública no es una tarea sencilla. En realidad, se han propuesto numerosos sistemas de este tipo en la literatura, aunque desafortunadamente la mayoría han demostrado ser inseguros. Hasta la fecha, la puesta en práctica más útil de la criptografía de clave pública es el *sistema Rivest-Shamir-Adleman*,[11] el cual utiliza ideas de la teoría de números clásica. Éste se considera uno de los sistemas criptográficos más seguros porque ha soportado muchos intentos de expertos en el campo para vulnerarlo.

El algoritmo RSA es un cifrado en bloque que se basa en el hecho de que es fácil computacionalmente determinar un número primo aleatorio de gran tamaño (por ejemplo, de cien dígitos), aunque en la actualidad se considera computacionalmente infactible factorizar el producto de dos de tales números. De manera precisa, el cálculo de los parámetros específicos del algoritmo RSA procede del siguiente modo:

1. Se eligen al azar dos *números primos muy grandes*, p y q; los números primos tienen que escogerse con bastante cuidado, ya que algunos números primos dan lugar a un sistema muy débil.
2. Se multiplican los números p y q y se obtiene el producto

$$pq = n \tag{A5.30}$$

Se encuentra la *función indicada de Euler* de n, utilizando la fórmula

$$\phi(n) = (p-1)(q-1) \tag{A5.31}$$

La ecuación (A5.31) se deriva de la definición de la función indicada de Euler $\phi(n)$ como el número de enteros positivos i menores que n, tales que el *máximo común divisor* de i y n es igual a uno.

3. Sea e un entero positivo menor que $\phi(n)$, tal que el máximo común divisor de e y $\phi(n)$ es igual a uno. Por tanto, se encuentra un entero positivo d menor que $\phi(n)$, tal que

$$de = 1 \bmod \phi(n) \tag{A5.32}$$

La función de trampa de una vía RSA se define entonces calculando simplemente la *exponenciación discreta*

$$f_z(x) = x^e = y \bmod n \tag{A5.33}$$

Los valores de n y e constituyen la clave pública; de ahí que hacer público el algoritmo de fácil determinación E_z para calcular la función f_z equivale exactamente a dar a conocer los número n y e.

Los números primos p y q constituyen la clave privada. Puesto que d se relaciona con p y q, la sucesión del algoritmo D_z fácil de determinar (cuando se conoce la trampa z) para calcular la función inversa f_z^{-1} equivale exactamente a conocer p y q. En particular, la función inversa está definida por

$$f_z^{-1}(y) = y^d \bmod n \tag{A5.34}$$

El exponente de desencriptado d se encuentra utilizando la ecuación (A5.32), la cual es equivalente al enunciado (en la aritmética de enteros ordinaria) de que

$$de = \phi(n)Q + 1 \tag{A5.35}$$

para algún entero Q. Advierta que $\phi(n)$ se relaciona a su vez con p y q por medio de la ecuación (A5.31). Puesto que $y = x^e$, es posible utilizar las ecuaciones (A5.32) y (A5.33) para escribir

$$\begin{aligned} y^d &= x^{de} \\ &= x^{\phi(n)Q+1} \\ &= ((x^{\phi(n)})^Q)x \end{aligned} \tag{A5.36}$$

Aprovechamos ahora el celebrado teorema de Euler, el cual establece que para cualesquiera enteros positivos x y n con $x < n$, tenemos

$$x^{\phi(n)} = 1 \bmod n \tag{A5.37}$$

Por consiguiente, el uso de la ecuación (A5.37) en la (A5.36) produce el desencriptado que se desea

$$y^d = x \tag{A5.38}$$

De este modo vemos que la determinación de la función inversa f_z^{-1} es fácil, dado el conocimiento de los números primos p y q.

La *seguridad* del criptoalgoritmo RSA se sustenta en la premisa de que cualquier método de invertir la función f_z es *equivalente a factorizar* $n = pq$. Esta equivalencia plantea la pregunta: ¿es factible computacionalmente un ataque al factorizar n? Parece que la respuesta es no, siempre que los números primos p y q sean cada uno del orden de 100 dígitos decimales y que no haya un avance revolucionario en los algoritmos de factorización.

■ FIRMAS DIGITALES[12]

Para que un sistema de correo electrónico sustituya el uso del correo en papel ordinario en las transacciones comerciales, el usuario del sistema debe tener la posibilidad de "firmar" un men-

saje electrónico. El uso de una *firma digital* ofrece una prueba de que el mensaje lo originó el emisor. Para satisfacer este requerimiento, la firma digital debe tener las siguientes propiedades:

- El receptor de un mensaje electrónico es capaz de verificar la firma del emisor.
- La firma no es falsificable.
- El emisor de un mensaje electrónico firmado no puede negar que lo es.

Para poner en práctica firmas digitales utilizando el algoritmo RSA, procedemos de la manera siguiente. Un usuario que posee la clave privada d firma un bloque de mensaje dado m formando la firma

$$s = m^d \bmod n \tag{A5.39}$$

Resulta difícil calcular s a menos que se conozca la clave privada d. Por tanto, es difícil falsificar una firma digital definida de acuerdo con la ecuación (A5.39). Además, el emisor del mensaje m no puede negar haberlo enviado, ya que nadie más podía haber creado la firma s. El receptor procede utilizando la clave pública e para calcular

$$\begin{aligned} s^e &= (m^d)^e \bmod n \\ &= m^{de} \bmod n \\ &= m \bmod n \end{aligned} \tag{A5.40}$$

donde, en la última línea se usó la ecuación (A5.32). De ese modo, el receptor es capaz de validar la firma del emisor estableciendo que el cálculo de $s^e \bmod n$ produce el mismo resultado que el mensaje descifrado m. El algoritmo RSA satisface de ese modo las tres propiedades necesarias de una firma digital.

A5.7 *Resumen y discusión*

La criptografía es un área de investigación "viva". Esta afirmación no debe sorprender. Considerando el hecho de que nos encontramos en una *sociedad de la información*, continuará creciendo la importancia de la criptografía como un mecanismo de seguridad. En este apéndice hemos presentado un tratamiento introductorio de este tema altamente importante.

Podemos clasificar la criptografía en la de clave secreta y en la de clave pública, dependiendo de si la clave que se usa para el encriptamiento de un mensaje y su desencriptamiento es por completo secreta o parcialmente pública. De modo alternativo, es factible clasificar un sistema criptográfico en uno cifrado de bloque o en uno de secuencia, dependiendo del método de puesta en práctica. Un cifrado de bloque presenta propagación de errores, lo cual quizás muestre ser altamente valioso en la autentificación.

Entre los muchos sistemas criptográficos creados hasta la fecha, el estándar de inscripción de datos (DES) y los algoritmos de Rivest-Shamir-Adleman (RSA) representan los de más éxito. Ambos criptoalgoritmos son cifrados de bloque. Difieren entre sí en que el algoritmo DES implica el uso de una clave secreta, en tanto que el RSA incluye el uso de una pública. En un sistema de clave secreta, tanto el emisor como el receptor comparten la misma clave. Por otro lado, en un sistema de clave pública, ésta se divide en dos partes. Una clave pública ubicada en el transmisor y una clave privada (secreta) situada en el receptor; en el último sistema, resulta infactible computacionalmente recuperar el mensaje del texto común a partir de su versión encriptada sin el conocimiento de la clave privada.

Aunque los criptosistemas de clave pública como el RSA proporcionan un método efectivo para la administración de claves, son ineficientes para el encriptado en masa de datos debido a los bajos anchos de banda. En contraste, los criptosistemas convencionales como el DES

proporcionan mejor rendimiento, pero requieren administración de claves. Esto sugiere el posible uso de un método híbrido que explote los mejores elementos de ambos sistemas como la base para el diseño práctico de un sistema de comunicación seguro. Por ejemplo, el algoritmo RSA se utilice tal vez para la autentificación y el algoritmo DES para el encriptado.

NOTAS Y BIBLIOGRAFÍA

1. Para un tratamiento introductorio de la criptografía, vea el capítulo 15 del libro de Adamék [3]. En cuanto a un tratamiento amplio de muchas facetas de la criptología, consulte el libro que edita Simmons [138]; este libro es una edición ampliada de un número especial de los *Proceedings del IEEE* sobre criptología. Las aportaciones del capítulo del libro de Simmons las escriben autoridades que van a la vanguardia en el tema de la criptología. En el libro de van Tilborg [159] se presenta también un tratamiento ameno de esta materia.
2. La era de la criptografía científica de clave secreta se anunció con la publicación del memorable artículo de Shannon [370], el cual estableció la conexión entre la criptografía y la teoría de la información.
3. La era de la criptografía de clave pública se estableció con la publicación de otro artículo memorable de Diffie y Hellman [233], quienes demostraron por primera vez que es posible tener una comunicación secreta sin ninguna transferencia de clave entre el emisor y el receptor. Fue el artículo de Diffie y Hellman el que detonó el interés por la investigación en la criptología, mismo que ha continuado desde entonces.
4. El término *criptoanalista enemigo* suele usarse en la criptología para referirse a un interceptor (escuchador furtivo) de criptogramas; su uso se originó de las aplicaciones militares.
5. Con respecto a un tratamiento amplio de los cifrados de secuencia, consulte el capítulo 2 escrito por R.A. Rueppel en el libro *Contemporary Cryptology*, editado por Simmons [138].
6. En lo referente a una explicación altamente legible del modelo de criptografía de Shannon, vea el capítulo de introducción de J.L. Massey editado por Simmons [138].
7. Los blocs de un tiempo derivan su nombre del uso (un poco antes, durante y después de la Segunda Guerra Mundial) que de ellos hacían espías de varios gobiernos, a quienes se les entregaba un bloc de papel con una clave elegida al azar y se les indicaba utilizarla sólo para un encriptado sencillo. El bloc de un tiempo también se conoce como cifrado de Vernam, denominado así en reconocimiento de su creador, G.S. Vernam.
8. En cuanto a una deducción de la ecuación (A5.11), vea el artículo original de Shannon [370, 371].
9. La historia del algoritmo DES la narran M.E. Smid y D.K. Branstad en el capítulo 1 del libro editado por Simmons [138]. En relación con una descripción del algoritmo DES, vea Diffie y Hellman [234]. Consulte también los libros de Meyer y Matyas [106] y Torrieri ([155], capítulo 6).
10. En lo que respecta a un tratamiento amplio de la criptografía de clave pública, vea el capítulo 4 de J. Nechvatal en el libro editado por Simmons [138]. Este libro incluye también una contribución de capítulos de W. Diffie que describe los diversos intentos para idear criptoalgoritmos seguros de clave pública y la evolución gradual de una variedad de protocolos basados en ellos.
11. El sistema RSA está patentado; recibe su nombre debido a sus creadores R. L. Rivest, A. Shamir y L. Adleman. La referencia original de este criptosistema es Rivest, Shamir y Adleman [356].
12. La idea de una firma digital fue analizada por primera vez por Diffie y Hellman [233]. Su puesta en práctica utilizando el algoritmo RSA se describe en Rivest, Shamir y Adleman [356]. En cuanto a un tratamiento detallado de las firmas digitales, vea el capítulo 6 de C.J. Michell, F. Piper y R. Wild en el libro editado por Simmons [138].

APÉNDICE 6

TABLAS

Las doce tablas que se compilan en este apéndice final incluyen:

- *Código ASCII*
- *Transformadas de Fourier y de Hilbert*
- *Funciones de Bessel*
- *Función de error*
- *Estándares de módem seleccionados*
- *Identidades trigonométricas, expansiones en series e integrales*
- *Constantes útiles y prefijos de unidades recomendados*

TABLA A6.1 *Código ASCII*

Posición de bit											
				7	*0*	*0*	*0*	*0*	*1*	*1*	*1*	*1*
				6	*0*	*0*	*0*	*1*	*1*	*0*	*1*	*1*
4	*3*	*2*	*1*	*5*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*
0	0	0	0	NUL		DLE	SP	0	@	P	\	p
0	0	0	1	SOH		DC1	!	1	A	Q	a	q
0	0	1	0	STX		DC2	"	2	B	R	b	r
0	0	1	1	ETX		DC3	#	3	C	S	c	s
0	1	0	0	EOT		DC4	$	4	D	T	d	t
0	1	0	1	ENQ		NAK	%	5	E	U	e	u
0	1	1	0	ACK		SYN	&	6	F	V	f	v
0	1	1	1	BEL		ETB	'	7	G	W	g	w
1	0	0	0	BS		CAN	(	8	H	X	h	x
1	0	0	1	HT		EM	)	9	I	Y	i	y
1	0	1	0	LF		SUB	*	:	J	Z	j	z
1	0	1	1	VT		ESC	+	;	K	[	k	{
1	1	0	0	FF		FS	'	<	L	\	l	:
1	1	0	1	CR		GS	–	=	M	]	m	}
1	1	1	0	SO		RS	.	>	N	^	n	~
1	1	1	1	SI		US	/	?	O	—	o	DEL

ACK	Reconocimiento	ENQ	Preguntar	NUL	Nulo o puros ceros
BEL	Campana o alarma	EOT	Fin de transmisión	RS	Separador de registro
BS	Retroceso	ESC	Escape	SI	Activar mayúscula
CAN	Cancelar	ETB	Fin del bloque de transmisión	SO	Desactivar mayúscula
CR	Retorno de carro			SOH	Inicio o encabezado
DC1	Control de dispositivo 1	ETX	Fin del texto	SP	Espacio
DC2	Control de dispositivo 2	FF	Alimentación de formas	STX	Inicio del texto
DC3	Control de dispositivo 3	FS	Separador de archivos	SUB	Sustituto
DC4	Control de dispositivo 4	GS	Separador de grupo	SYN	Desocupamiento síncrono
DEL	Borrar	HT	Tabulador horizontal		
DLE	Escape del enlace de datos	LF	Alimentación de línea	US	Separador de unidad
EM	Fin del medio	NAK	Reconocimiento negativo	VT	Tabulador vertical

(Tomado de Couch, con autorización de Macmillan.)

TABLA A6.2 *Resumen de propiedades de la transformada de Fourier*

Propiedad	*Descripción matemática*
1. Linealidad	$ag_1(t) + bg_2(t) \rightleftharpoons aG_1(f) + bG_2(f)$ donde a y b son constantes
2. Escalamiento de tiempo	$g(at) \rightleftharpoons \frac{1}{\lvert a \rvert} G\left(\frac{f}{a}\right)$ donde a es una constante
3. Dualidad	Si $g(t) \rightleftharpoons G(f)$, entonces $G(t) \rightleftharpoons g(-f)$
4. Corrimiento en el tiempo	$g(t - t_0) \rightleftharpoons G(f) \exp(-j2\pi f t_0)$
5. Corrimiento en frecuencia	$\exp(j2\pi f_c t) g(t) \rightleftharpoons G(f - f_c)$
6. Área bajo $g(t)$	$\int_{-\infty}^{\infty} g(t)\, dt = G(0)$
7. Área bajo $G(f)$	$g(0) = \int_{-\infty}^{\infty} G(f)\, df$
8. Diferenciación en el dominio del tiempo	$\frac{d}{dt} g(t) \rightleftharpoons j2\pi f G(f)$
9. Integración en el dominio del tiempo	$\int_{-\infty}^{t} g(\tau)\, d\tau \rightleftharpoons \frac{1}{j2\pi f} G(f) + \frac{G(0)}{2} \delta(f)$
10. Funciones conjugadas	Si $g(t) \rightleftharpoons G(f)$, entonces $g^*(t) \rightleftharpoons G^*(-f)$
11. Multiplicación en el dominio del tiempo	$g_1(t) g_2(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\lambda) G_2(f - \lambda)\, d\lambda$
12. Convolución en el dominio del tiempo	$\int_{-\infty}^{\infty} g_1(\tau) g_2(t - \tau)\, d\tau \rightleftharpoons G_1(f) G_2(f)$

TABLA A6.3 *Pares de transformadas de Fourier*

Función de tiempo	*Transformada de Fourier*
$\text{rect}\left(\frac{t}{T}\right)$	$T\,\text{senc}(fT)$
$\text{senc}(2Wt)$	$\frac{1}{2W}\text{rect}\left(\frac{f}{2W}\right)$
$\exp(-at)u(t), \quad a>0$	$\frac{1}{a+j2\pi f}$
$\exp(-a\lvert t\rvert), \quad a>0$	$\frac{2a}{a^2+(2\pi f)^2}$
$\exp(-\pi t^2)$	$\exp(-\pi f^2)$
$\begin{cases} 1-\frac{\lvert t\rvert}{T}, & \lvert t\rvert < T \\ 0, & \lvert t\rvert \geq T \end{cases}$	$T\,\text{senc}^2(fT)$
$\delta(t)$	1
1	$\delta(f)$
$\delta(f-t_0)$	$\exp(-j2\pi f t_0)$
$\exp(j2\pi f_c t)$	$\delta(f-f_c)$
$\cos(2\pi f_c t)$	$\frac{1}{2}[\delta(f-f_c)+\delta(f+f_c)]$
$\text{sen}(2\pi f_c t)$	$\frac{1}{2j}[\delta(f-f_c)-\delta(f+f_c)]$
$\text{sgn}(t)$	$\frac{1}{j\pi f}$
$\frac{1}{\pi t}$	$-j\,\text{sgn}(f)$
$u(t)$	$\frac{1}{2}\delta(f)+\frac{1}{j2\pi f}$
$\sum_{i=-\infty}^{\infty}\delta(t-iT_0)$	$\frac{1}{T_0}\sum_{n=-\infty}^{\infty}\delta\left(f-\frac{n}{T_0}\right)$

Notas: $u(t)$ = función escalón unitario
$\delta(t)$ = función delta o impulso unitario
$\text{rec}(t)$ = función rectangular de amplitud unitaria y duración unitaria centrada en el origen
$\text{sgn}(t)$ = función signo
$\text{senc}(t)$ = función senc

TABLA A6.4 *Pares de transformadas de Hilbert*[a]

Función de tiempo	*Transformada de Hilbert*
$m(t)\cos(2\pi f_c t)$	$m(t)\,\mathrm{sen}(2\pi f_c t)$
$m(t)\,\mathrm{sen}(2\pi f_c t)$	$-m(t)\cos(2\pi f_c t)$
$\cos(2\pi f_c t)$	$\mathrm{sen}(2\pi f_c t)$
$\mathrm{sen}(2\pi f_c t)$	$-\cos(2\pi f_c t)$
$\dfrac{\mathrm{sen}\, t}{t}$	$\dfrac{1-\cos t}{t}$
$\mathrm{rect}(t)$	$-\dfrac{1}{\pi}\log\left\lvert\dfrac{t-\frac{1}{2}}{t+\frac{1}{2}}\right\rvert$
$\delta(t)$	$\dfrac{1}{\pi t}$
$\dfrac{1}{1+t^2}$	$\dfrac{t}{1+t^2}$
$\dfrac{1}{t}$	$-\pi\delta(t)$

[a] En los primeros dos pares, se supone que $m(t)$ está limitada en banda al intervalo $-W \le f \le W$, donde $W < f_c$.

Notas: $\delta(t)$: función delta
rect(t): función rectangular de amplitud unitaria y duración unitaria centrada en el origen
log: logaritmo natural

TABLA A6.5 *Tabla de funciones de Bessel*[a]

	$J_n(x)$								
$n\backslash x$	0.5	1	2	3	4	6	8	10	12
0	0.9385	0.7652	0.2239	−0.2601	−0.3971	0.1506	0.1717	−0.2459	0.0477
1	0.2423	0.4401	0.5767	0.3391	−0.0660	−0.2767	0.2346	0.0435	−0.2234
2	0.0306	0.1149	0.3528	0.4861	0.3641	−0.2429	−0.1130	0.2546	−0.0849
3	0.0026	0.0196	0.1289	0.3091	0.4302	0.1148	−0.2911	0.0584	0.1951
4	0.0002	0.0025	0.0340	0.1320	0.2811	0.3576	−0.1054	−0.2196	0.1825
5	–	0.0002	0.0070	0.0430	0.1321	0.3621	0.1858	−0.2341	−0.0735
6		–	0.0012	0.0114	0.0491	0.2458	0.3376	−0.0145	−0.2437
7			0.0002	0.0025	0.0152	0.1296	0.3206	0.2167	−0.1703
8			–	0.0005	0.0040	0.0565	0.2235	0.3179	0.0451
9				0.0001	0.0009	0.0212	0.1263	0.2919	0.2304
10					0.0002	0.0070	0.0608	0.2075	0.3005
11					–	0.0020	0.0256	0.1231	0.2704
12						0.0005	0.0096	0.0634	0.1953
13						0.0001	0.0033	0.0290	0.1201
14						–	0.0010	0.0120	0.0650

[a] Para tablas más amplias de la función de Bessel, vea Watson ([163], pp. 666-697) y Abramowitz y Stegun ([1], pp. 358-406)

TABLA A6.6 *La función de error*[a]

u	$fer(u)$	u	$fer(u)$
0.00	0.00000	1.10	0.88021
0.05	0.05637	1.15	0.89612
0.10	0.11246	1.20	0.91031
0.15	0.16800	1.25	0.92290
0.20	0.22270	1.30	0.93401
0.25	0.27633	1.35	0.94376
0.30	0.32863	1.40	0.95229
0.35	0.37938	1.45	0.95970
0.40	0.42839	1.50	0.96611
0.45	0.47548	1.55	0.97162
0.50	0.52050	1.60	0.97635
0.55	0.56332	1.65	0.98038
0.60	0.60386	1.70	0.98379
0.65	0.64203	1.75	0.98667
0.70	0.67780	1.80	0.98909
0.75	0.71116	1.85	0.99111
0.80	0.74210	1.90	0.99279
0.85	0.77067	1.95	0.99418
0.90	0.79691	2.00	0.99532
0.95	0.82089	2.50	0.99959
1.00	0.84270	3.00	0.99998
1.05	0.86244	3.30	0.999998

[a]La función de error se tabula extensivamente en varias referencias; vea, por ejemplo, Abramowitz y Stegun ([1], pp. 297-316).

PROYECTO UNI / ASDI (SAREC) / FEC
UNIVERSIDAD NACIONAL DE INGENIERIA
FACULTAD DE ELECTROTECNIA Y COMPUTACION

TABLA A6.7 *Selección de estándares de módem de banda de voz (línea telefónica) ITU*

	Estándar[a] *ITU*	*Tipo de modulación*	*Tasa de bits/s*	*Tasa de símbolo, bauds*
a) Módems simétricos:	V.21	FSK binaria	300	300
	V.22 bis	QPSK	1200	600
	V.26	QPSK	2400	1200
	V.27	PSK 8	4800	2400
	V.32	QAM 16	9600	2400
	V.34	QAM 1024	28,800	3429
	V.34 Alta velocidad	Constelación anidada de cuatro constelaciones QAM 960	33,600	
b) Módems asimétricos	V.90: Sentido directo	Digital	56,000	
	Sentido inverso	Alta velocidad	33,600	

[a]El sufijo "bis" designa la segunda versión de un estándar particular.

TABLA A6.8 *Identidades trigonométricas*

$$\exp(\pm j\theta) = \cos\theta \pm j\,\text{sen}\,\theta$$

$$\cos\theta = \tfrac{1}{2}[\exp(j\theta) + \exp(-j\theta)]$$

$$\text{sen}\,\theta = \frac{1}{2j}[\exp(j\theta) - \exp(-j\theta)]$$

$$\text{sen}^2\theta + \cos^2\theta = 1$$

$$\cos^2\theta - \text{sen}^2\theta = \cos(2\theta)$$

$$\cos^2\theta = \tfrac{1}{2}[1 + \cos(2\theta)]$$

$$\text{sen}^2\theta = \tfrac{1}{2}[1 - \cos(2\theta)]$$

$$2\,\text{sen}\,\theta\cos\theta = \text{sen}(2\theta)$$

$$\text{sen}(\alpha \pm \beta) = \text{sen}\,\alpha\cos\beta \pm \cos\alpha\,\text{sen}\,\beta$$

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \pm \text{sen}\,\alpha\,\text{sen}\,\beta$$

$$\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$$

$$\text{sen}\,\alpha\,\text{sen}\,\beta = \tfrac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta)]$$

$$\cos\alpha\cos\beta = \tfrac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$$

$$\text{sen}\,\alpha\cos\beta = \tfrac{1}{2}[\text{sen}(\alpha - \beta) - \text{sen}(\alpha + \beta)]$$

TABLA A6.9 *Expansiones en serie*

Serie de Taylor

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n+\cdots$$

donde

$$f^{(n)}(a)=\left.\frac{d^n f(x)}{dx^n}\right|_{x=a}$$

Serie de MacLaurin

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\cdots+\frac{f^{(n)}(0)}{n!}x^n+\cdots$$

donde

$$f^{(n)}(0)=\left.\frac{d^n f(x)}{dx^n}\right|_{x=0}$$

Serie del binomio

$$(1+x)^n=1+nx+\frac{n(n-1)}{2!}x^2+\cdots, \qquad |nx|<1$$

Serie exponencial

$$\exp x=1+x+\frac{1}{2!}x^2+\cdots$$

Serie logarítmica

$$\log(1+x)=x-\tfrac{1}{2}x^2+\tfrac{1}{3}x^3-\cdots$$

Serie trigonométrica

$$\operatorname{sen} x=x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\cdots$$

$$\cos x=1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\cdots$$

$$\tan x=x+\frac{1}{3}x^3+\frac{1}{15}x^5+\cdots$$

$$\operatorname{sen}^{-1} x=x+\frac{1}{6}x^3+\frac{3}{40}x^5+\cdots$$

$$\tan^{-1} x=x-\frac{1}{3}x^3+\frac{1}{5}x^5-\cdots, \qquad |x|<1$$

$$\operatorname{senc} x=1-\frac{1}{3!}(\pi x)^2+\frac{1}{5!}(\pi x)^4-\cdots$$

TABLA A6.10 *Integrales*

Integrales indefinidas

$$\int x \operatorname{sen}(ax)\,dx = \frac{1}{a^2}[\operatorname{sen}(ax) - ax\cos(ax)]$$

$$\int x\cos(ax)\,dx = \frac{1}{a^2}[\cos(ax) - ax\operatorname{sen}(ax)]$$

$$\int x\exp(ax)\,dx = \frac{1}{a^2}\exp(ax)(ax-1)$$

$$\int x\exp(ax^2)\,dx = \frac{1}{2a}\exp(ax)^2$$

$$\int \exp(ax)\operatorname{sen}(bx)\,dx = \frac{1}{a^2+b^2}\exp(ax)[a\operatorname{sen}(bx) - b\cos(bx)]$$

$$\int \exp(ax)\cos(bx)\,dx = \frac{1}{a^2+b^2}\exp(ax)[a\cos(bx) - b\operatorname{sen}(bx)]$$

$$\int \frac{dx}{a^2+b^2x^2} = \frac{1}{ab}\tan^{-1}\left(\frac{bx}{a}\right)$$

$$\int \frac{x^2\,dx}{a^2+b^2x^2} = \frac{x}{b^2} - \frac{a}{b^3}\tan^{-1}\left(\frac{bx}{a}\right)$$

Integrales definidas

$$\int_0^\infty \frac{x\operatorname{sen}(ax)}{b^2+x^2}\,dx = \frac{\pi}{2}\exp(-ab), \qquad a>0,\ b>0$$

$$\int_0^\infty \frac{\cos(ax)}{b^2+x^2}\,dx = \frac{\pi}{2b}\exp(-ab), \qquad a>0,\ b>0$$

$$\int_0^\infty \frac{\cos(ax)}{(b^2-x^2)^2}\,dx = \frac{\pi}{4b^3}[\operatorname{sen}(ab) - ab\cos(ab)], \qquad a>0,\ b>0$$

$$\int_0^\infty \operatorname{senc} x\,dx = \int_0^\infty \operatorname{senc}^2 x\,dx = \frac{1}{2}$$

$$\int_0^\infty \exp(-ax^2)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}, \qquad a>0$$

$$\int_0^\infty x^2\exp(-ax^2)\,dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}}, \qquad a>0$$

TABLA A6.11 *Constantes útiles*

Constantes físicas	
Constante de Boltzmann	$k = 1.38 \times 10^{-23}$ joule/grados Kelvin
Constante de Planck	$h = 6.626 \times 10^{-34}$ joule-segundo
Carga del electrón (fundamental)	$q = 1.602 \times 10^{-19}$ coulomb
Velocidad de la luz en el vacío	$c = 2.998 \times 10^{8}$ metros/segundo
Temperatura estándar (absoluta)	$T_0 = 273$ grados Kelvin
Voltaje térmico	$V_T = 0.026$ volt a temperatura ambiente
Energía térmica kT a la temperatura estándar	$kT_0 = 3.77 \times 10^{-21}$ joule
Un hertz (hz) = 1 ciclo/segundo; 1 ciclo = 2π radianes	
Un watt (W) =1 joule/segundo	
Constantes matemáticas	
Base del logaritmo natural	$e = 2.7182818$
Logaritmo de e base 2	$\log_2 e = 1.442695$
Logaritmo de 2 base e	$\log 2 = 0.693147$
Logaritmo de 2 base 10	$\log_{10} 2 = 0.30103$
Pi	$\pi = 3.1415927$

TABLA A6.12 *Prefijos de unidades recomendados*

Múltiplos y submúltiplos	*Prefijos*	*Símbolo*
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	K (*k*)
10^{-3}	mili	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p

GLOSARIO

Convenciones y notaciones

1. El símbolo | | significa valor absoluto, o magnitud, de la cantidad compleja contenida dentro de él.
2. El símbolo arg( ) significa el ángulo de fase de la cantidad compleja contenida dentro de él.
3. El símbolo Re[ ] indica la "parte real de", e Im[ ] señala la "parte imaginaria de".
4. A menos que se establezca de otro modo, el algoritmo natural se denota por medio de log. Los logaritmos de las bases 2 y 10 se denotan mediante $\log_2$ y $\log_{10}$, respectivamente.
5. El uso de un asterisco como exponente denota conjugada compleja, es decir, x^* es la conjugada compleja de x.
6. El símbolo $\rightleftharpoons$ indica un par de transformadas de Fourier, por ejemplo, $g(t) \rightleftharpoons G(f)$, donde una letra minúscula denota la función de tiempo y la correspondiente mayúscula indica la función de frecuencia.
7. El símbolo $F[\ ]$ indica la operación de la transformada de Fourier, es decir, $F[g(t)] = G(f)$, y el símbolo $F^{-1}[\ ]$ indica la operación de la transformada inversa de Fourier, es decir, $F^{-1}[G(f)] = g(t)$.
8. El símbolo $\bigstar$ denota convolución, es decir,

$$x(t) \bigstar h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,d\tau$$

9. El símbolo $\oplus$ denota suma módulo 2, salvo en el capítulo 10, donde se usa la aritmética binaria y la suma módulo 2 se denota mediante un signo más ordinario.
10. El uso del subíndice T_0 indica que la función pertinente $g_{T_0}(t)$, digamos, es una función periódica del tiempo t con periodo T_0.
11. El uso de un sombrero sobre una función indica una de dos cosas:
    a) La transformada de Hilbert de una función; por ejemplo, la función $\hat{g}(t)$ es la transformada de Hilbert de $g(t)$, o
    b) La estimación de un parámetro desconocido; por ejemplo, la cantidad $\hat{\alpha}(\mathbf{x})$ es una estimación del parámetro desconocido α, con base en el vector de observación $\mathbf{x}$.
12. El uso de una tilde en una función indica la envolvente compleja de una señal de banda angosta; por ejemplo, la función $\tilde{g}(t)$ es la envolvente compleja de la señal de banda angosta $g(t)$. La excepción para esta convención está en la sección 10.8, donde, en la descripción de la decodificación de turbo, la tilde se usa para dar a entender información intrínseca y distinguirla de ese modo de la proporción de verosimilitud logarítmica.
13. El uso del subíndice + indica la preenvolvente de una señal; por ejemplo, la función $g_+(t)$ es la envolvente de la señal $g(t)$. De esa manera escribimos $g_+(t) = g(t) + j\hat{g}(t)$, donde $\hat{g}(t)$ es la transformada de Hilbert de $g(t)$. El uso del subíndice – indica que $g_-(t) = g(t) - j\hat{g}(t) = g_+^*(t)$.
14. El uso de los subíndices I y Q indica las componentes en fase y en cuadratura de una señal de banda angosta, un proceso aleatorio de banda angosta o la respuesta al impulso de un filtro de banda angosta, con respecto a la portadora $\cos(2\pi f_c t)$.

15. En una señal de mensaje pasobajas, la componente de frecuencia más alta o el ancho de banda del mensaje se denotan por medio de W. El espectro de esta señal ocupa el intervalo de frecuencia $-W \leq f \leq W$ y es cero en otro lado. En una señal pasobanda con frecuencia f_c de portadora, el espectro ocupa los intervalos de frecuencia $f_c - W \leq f \leq f_c + W$ y $-f_c - W \leq f \leq -f_c + W$, y de esa manera $2W$ denota el ancho de banda de la señal. La envolvente compleja (pasobajas) de esta señal pasobanda tiene un espectro que ocupa el intervalo de frecuencia $-W \leq f \leq W$.

    En el caso de un filtro pasobajas, el ancho de banda se denota mediante B. Una definición común del ancho de banda del filtro es la frecuencia a la cual la respuesta en magnitud del filtro disminuye en 3 dB por debajo del valor de la frecuencia cero. Para un filtro pasobanda de frecuencia de banda media f_c, el ancho de banda se denota mediante $2B$, centrado en f_c. El equivalente pasobajas complejo de este filtro pasobajas tiene un ancho de banda igual a B.

    El ancho de banda de transmisión de un canal de comunicación, requerido para transmitir una onda modulada, se denota por medio de B_T.
16. Las variables aleatorias o vectores aleatorios son letras mayúsculas (por ejemplo, X o $\mathbf{X}$) y sus valores de muestra son letras minúsculas (digamos, x o $\mathbf{x}$).
17. Una barra vertical en una expresión significa "dado que"; por ejemplo, $f_X(x|H_0)$ es la función de densidad de probabilidad de la variable aleatoria X, dado que la hipótesis H_0 es cierta.
18. El símbolo $E[\ ]$ significa el valor esperado de la variable aleatoria encerrado dentro de él. La E actúa como un operador.
19. El símbolo var[ ] indica la varianza de la variable aleatoria encerrada dentro de los corchetes.
20. El símbolo cov[ ] significa la covarianza de las dos variables aleatorias encerradas dentro de los corchetes.
21. La probabilidad promedio del error de símbolo se denota por medio de P_e.

    En el caso de técnicas de transmisión de señales binarias, p_{10} denota la probabilidad condicional del error, dado que el símbolo 0 se transmitió, y p_{01} denota la probabilidad condicional del error, dado que se transmitió el símbolo 1. Las probabilidades *a priori* de los símbolos 0 y 1 se denotan mediante p_0 y p_1, respectivamente.
22. El símbolo $\langle\ \rangle$ denota el promedio en el tiempo de la función de muestra encerrada dentro de él.
23. La letra negrita denota un vector o matriz. La inversa de una matriz cuadrada $\mathbf{R}$ se indica mediante $\mathbf{R}^{-1}$. La transpuesta de un vector $\mathbf{w}$ se denota mediante $\mathbf{w}^T$. La transpuesta hermitiana de un vector $\mathbf{x}$ de valores complejos se indica por medio de $\mathbf{x}^H$; la transposición hermitiana indica tanto transposición como conjugación compleja.
24. La longitud de un vector $\mathbf{x}$ se denota por medio de $||\mathbf{x}||$. La distancia euclidiana entre los vectores $\mathbf{x}_i$ y $\mathbf{x}_j$ se indica mediante $d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$.
25. El producto interior de dos vectores $\mathbf{X}$ y $\mathbf{Y}$ de valores reales se denota por medio de $\mathbf{x}^T\mathbf{y}$; su producto exterior se denota por medio de $\mathbf{x}\mathbf{y}^T$. Si los vectores $\mathbf{x}$ y $\mathbf{y}$ son de valor complejo, su producto interior es $\mathbf{x}^H\mathbf{y}$, y su productor exterior corresponde a $\mathbf{x}\mathbf{y}^H$.
26. El producto vectorial de dos vectores $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$ de M por 1 es un vector de M por 1 definido por

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \begin{bmatrix} \alpha_1\beta_1 \\ \alpha_1\beta_1 \\ \vdots \\ \alpha_M\beta_M \end{bmatrix}$$

donde α_k y β_k son los k-ésimos elementos de $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$, respectivamente. La norma L_1 del producto vectorial $\boldsymbol{\alpha} \cdot \boldsymbol{\beta}$ está definida por

$$\| \boldsymbol{\alpha} \cdot \boldsymbol{\beta} \|_1 = \sum_{m=1}^{M} \alpha_m\beta_m$$

Funciones

1. Función rectangular: $\text{rect}(t) = \begin{cases} 1, & -\frac{1}{2} < t < \frac{1}{2} \\ 0, & |t| > \frac{1}{2} \end{cases}$

2. Función de escalón unitario: $u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases}$

3. Función signo: $\text{sgn}(t) = \begin{cases} 1, & t > 0 \\ 0, & t = 0 \\ -1, & t < 0 \end{cases}$

4. Función delta (Dirac): $\delta(t) = 0, \quad t \neq 0$

$$\int_{-\infty}^{\infty} \delta(t)\, dt = 1$$

o, equivalentemente,

$$\int_{-\infty}^{\infty} g(t)\, \delta(t - t_0)\, dt = g(t_0)$$

5. Función senc: $\text{senc}(x) = \dfrac{\text{sen}(\pi x)}{\pi x}$

6. Integral seno: $\text{Si}(u) = \int_0^u \dfrac{\text{sen}\, x}{x}\, dx$

7. Función de error: $\text{fer}(u) = \dfrac{2}{\sqrt{\pi}} \int_0^u \exp(-z^2)\, dz$

Función de error complementario: $\text{ferc}(u) = 1 - \text{fer}(u).$

8. Coeficiente del binomio $\dbinom{n}{k} = \dfrac{n!}{(n-k)!\,k!}$

9. Función de Bessel del primer tipo de orden n: $J_n(x) = \dfrac{1}{2\pi} \int_{-\pi}^{\pi} \exp(jx \text{ sen } \theta - jn\theta)\, d\theta$

10. Función de Bessel modificada del primer tipo de orden cero: $I_0(x) = \dfrac{1}{2\pi} \int_{-\pi}^{\pi} \exp(x \cos \theta)\, d\theta$

11. Función hipergeométrica confluente $${}_1F_1(a; b; x) = 1 + \frac{a}{b}\frac{x}{1!} + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \cdots$$

Abreviaturas

A: ampere
ADC: (analog-to-digital converter) convertidor analógico digital
ADM: (adaptive delta modulation) modulación delta adaptable
ADPCM: (adaptive differential pulse-code modulation) modulación adaptable diferencial por codificación de pulsos
ADSL: (asymmetric digital subscriber line) línea digital asimétrica de suscriptor
AM: (amplitude modulation) modulación de amplitud
ANSI: (American National Standards Institute) Instituto Estadounidense de Estándares Nacionales
APB: (adaptive prediction with backward estimation) predicción adaptable con estimación hacia atrás
APF: (adaptive prediction with forward estimation) predicción adaptable con estimación hacia adelante

AQB: (adaptive quantization with backward estimation) cuantización adaptable con estimación hacia atrás
AQF: (adaptive quantization with forward estimation) cuantización adaptable con estimación hacia adelante
ARQ: (automatic-repeat request) solicitud de repetición automática
ASCII: (American National Standard Code for Information Interchange) Código Estándar Estadounidense para el Intercambio de Información
ASK: (amplitude-shift keying) corrimiento de amplitud por llaveo
ATM: (asynchronous transfer mode) modo de transferencia asíncrona
AWGN: (additive white Gaussian noise) ruido blanco gaussiano aditivo
b/s: bits/segundo
BER: (bit error rate) tasa de error de bits
B-ISDN: (broadband ISDN) red digital de servicios integrados de banda ancha
BPF: (band-pass filter) filtro pasobanda
BSC: (binary symmetric channel) canal simétrico binario
ca: corriente alterna
CAP: (carrierless amplitude/phase modulation) modulación de amplitud/fase sin portadora
CCITT: (Consultative Committee for International Telephone and Telegraph) Comité Consultivo Internacional para la Telefonía y la Telegrafía (renombrada ahora ITU)
cd: corriente directa
CDM: (code-division multiplexing) multiplexado por división de código
CDMA: (code-division multiple access) acceso múltiple por división de código
CELP [code-excited linear predictive (model)] (modelo) predictivo lineal de código de excitación
codec: codificador/decodificador
CPFSK: (continuous-phase frequency-shift keying) corrimiento de frecuencia por llaveo en fase continua
CRC: (cyclic redundancy check) verificación cíclica de redundancia
CW: onda continua
DAC: (digital-to-analog converter) convertidor analógico-digital
dB: decibel
dBW: decibel referenciado a 1 watt
dBmW: decibel referenciado a 1 miliwatt
DEM: demodulador
DES: (data encryption standard) estándar de encriptado de datos
DFT: (discrete Fourier transform) transformada discreta de Fourier
DM: (delta modulation) modulación delta
DMT: (discrete multitone) multitonos discretos
DPCM: (differential pulse-code modulation) modulación diferencial por codificación de pulsos
DPSK: (differential phase-shift keying) corrimiento de fase por llaveo diferencial
DSB-SC: (double sideband-suppressed carrier) doble banda lateral con portadora suprimida
DS/BPSK: (direct sequence/binary phase-shift keying) corrimiento de fase por llaveo de secuencia directa/binaria
DSL: (digital subscriber line) línea de suscriptor digital
exp: exponencial
FDM: (frequency-division multiplexing) multiplexado por división de frecuencia

FDMA: (frequency-division multiple access) acceso múltiple por división de frecuencia
FEC: (feed-forward error correction) corrección de errores directa
FEXT: (far-end crosstalk) diafonía de extremo lejano
FFT: (fast Fourier transform) transformada rápida de Fourier
FH: (frequency hop) salto de frecuencia
FH/MFSK: (frequency hop/M-ary frequency-shift keying) salto de frecuencia/corrimiento de frecuencia por llaveo M-ario
FI: frecuencia intermedia
FMFB: (frequency modulator with feedback) modulador de frecuencia con realimentación
FSK: (frequencty-shift keying) corrimiento de frecuencia por llaveo
GMSK: (Gaussian filtered minimun shift keying) MSK de filtrado gaussiano
GSM: (global system for mobile communication) sistema global para comunicación móvil
HDTV: (high definition television) televisión de alta definición
Hz: Hertz
IDFT: (inverse discrete Fourier transform) transformada discreta de Fourier inversa
I/O: (input/output) entrada/salida
IP: (Internet protocol) protocolo de Internet
IS-95: (intermediate standard-95) estándar intermedio 95
ISDN: (integrated services digital network) red digital de servicios integrados
ISI: (intersymbol interference) interferencia de intersímbolo
ISO: (International Organization for Standardization) Organización Internacional para la Estandarización
ITU: (International Telecommunication Union) Unión Internacional de Telecomunicaciones
JPEG: (joint photographic experts group) Grupo Unido de Expertos en Fotografía
LAN: (local-area network) red de área local
LDM: (linear delta modulation) modulación delta lineal
LMS: (least-mean-square) mínimos cuadrados
log: logaritmo natural
$\log_2$: logaritmo en base 2
$\log_{10}$: logaritmo en base 10
LPC: [linear predictive coding (model)] (modelo) codificación predictiva lineal
LPF: (low-pass filter) filtro pasobajas
MAP: (maximum *a posteriori* probability) probabilidad máxima *a posteriori*
ML: (maximum likelihood) máxima verosimilitud
mmse: (minimum mean-square error) error cuadrático medio mínimo
módem: modulador-demodulador
MPEG: (motion photographic experts group) Grupo de Expertos de Imágenes en Movimiento
ms: milisegundo
μs: microsegundo
MSK: (minimum shift keying) corrimiento mínimo por llaveo
NCO: (number-controlled oscillator) oscilador controlado por número
NEXT: (near-end crosstalk) diafonía de extremo cercano
nm: nanómetro
NRZ: (nonreturn-to-zero) sin retorno a cero
NTSC: (National Television Systems Committee) Comité de Sistemas de Televisión Nacionales en Estados Unidos
OC: (optical carrier) portadora óptica

OFDM: (orthogonal frequency-division multiplexing) multiplexado por división de frecuencia ortogonal
OOK: (on-off keying) activado-desactivado por llaveo
OSI: (open systems interconnection) interconexión de sistemas abiertos
PAM: (pulse-amplitude modulation) modulación por amplitud de pulsos
PCM: (pulse-code modulation) modulación por codificación de pulsos
PDM: (pulse-duration modulation) modulación por duración de pulsos
PG: (processing gain) ganancia de procesamiento
PLL: (phase-locked loop) malla de fase encadenada
PN: (pseudo-noise) pseudo ruido
POTS: (plain old telephone service) servicio telefónico convencional
PPM: (pulse-position modulation) modulación por posición de pulsos
PSK: (phase-shift keying) corrimiento de fase por llaveo
PSTN: (public switched telephone network) red pública de conmutación telefónica
PWM: (pulse-width modulation) modulación por ancho de pulsos
QAM: (quadrature amplitude modulation) modulación en amplitud por cuadratura
QoS: (quality of service) calidad del servicios
QPSK: (quadriphase-shift keying) corrimiento de cuadrifase por llaveo
RF: radio frecuencia
rms: (root-mean-square) valor cuadrático medio
RS: Reed-Solomon
RS-232: [recommended standard-232 (port)] estándar recomendado 232 (puerto)
RSA: Rivest-Shamir-Adleman
RSC: [recursive systematic convolutional (code)] convolucional sistemático recursivo (código)
RZ: (return-to-cero) retorno a cero
s: segundo
SDH: (synchronous digital hierarchy) jerarquía digital síncrona
SDMA: (space-division multiple access) acceso múltiple por división de espacio
SDR: (signal-to-distortion ratio) relación señal a distorsión
SNR: (signal-to-noise ratio) relación señal a ruido
SONET: (synchronous optical network) red óptica síncrona
STFT: (short-time Fourier transform) transformada de Fourier de tiempo breve
STM: (synchronous transfer mode) modo de transferencia síncrona
TC: (time compression) compresión del tiempo
TCM: (trellis-coded modulation) modulación por código de enramado
TDM: (time-division multiplexing) multiplexado por división del tiempo
TDMA: (time-division multiple access) acceso múltiple por división del tiempo
TV: televisión
UHF: (ultra high frequency) ultra alta frecuencia
V: volt
VCO: (voltage-controlled oscillator) oscilador controlado por voltaje
VHF: (very high frequency) muy alta frecuencia
VLSI: (very-large-scale integration) integración a muy grande escala
W: watt
WDM: (wavelenght division multiplexing) multiplexado por división de longitud de onda

BIBLIOGRAFÍA

LIBROS

1. M. Abramowitz e I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Nueva York: Dover Publications).
2. N. Abramson, *Information Theory and Coding* (Nueva York: McGraw-Hill).
3. J. Adamék, *Foundations of Coding* (Nueva York: Wiley).
4. Y. Akaiwa, *Introduction to Digital Mobile Communication* (Nueva York: Wiley).
5. J.B. Anderson, T. Aulin y C.E. Sundberg, *Digital Phase Modulation* (Nueva York: Plenum Publishers).
6. J.B. Anderson y S. Mohan, *Source and Channel Coding: An Algorithmic Approach* (Boston, Mass: Kluwer Academic).
7. J.B. Anderson, *Digital Transmission Engineering* (Piscataway, NJ.: IEEE Press).
8. R.B. Ash, *Information Theory* (Nueva York: Wiley).
9. Bell Laboratories Technical Staff, *A History of Engineering Science in the Bell System: The Early Years (1875-1925)* (Libros sobre pedido, Ann Arbor, Michigan).
10. J.C. Bellamy, *Digital Telephony*, Segunda edición (Nueva York: Wiley).
11. S. Benedetto y E. Biglieri, *Principles of Digital Transmission with Wireless Applications* (Nueva York: Kluwer Academic/Plenum Publishers).
12. S. Benedetto, E. Biglieri y V. Castellani, *Digital Transmission Theory* (Englewood Cliffs, N.J.: Prentice-Hall).
13. W.R. Bennett, *Introduction to Signal Transmission* (Nueva York: McGraw-Hill).
14. K.B. Benson y J.C. Whitaker, *Television Engineering Handbook*, ed. rev. (Nueva York: McGrawHill).
15. T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression* (Englewood Cliffs, N.J.: Prentice-Hall).
16. E.R. Berlekamp, *Algebraic Coding Theory* (Nueva York: McGraw-Hill).
17. E.R. Berlekamp (editor), *Key Papers in the Development of Coding Theory* (Piscataway, N.J.: IEEE Press).
18. V.K. Bhargava, D. Haccoun, R. Matyas y P. Nuspl, *Digital Communications by Satellite: Modulation, Multiple Access, and Coding* (Nueva York: Wiley).
19. E. Biglieri, D. Divsalar, P.J. McLane y M.K. Simon, *Introduction to Trellis-Coded Modulation with Applications* (Nueva York: Macmillan).
20. J.A.C. Bingham, *The Theory and Practice of Modem Design* (Nueva York: Wiley).
21. R.B. Blachman y J.W. Tukey, *The Measurement of Power Spectra, from the Point of View of Communication Engineering* (Nueva York: Dover).
22. H.S. Black, *Modulation Theory* (Princeton, N.J.: Van Nostrand).
23. R.E. Blahut, *Principles y Practice of Information Theory* (Reading, Mass.: Addison-Wesley).
24. R.E. Blahut, *Digital Transmission of Infomation* (Reading, Mass.: Addison-Wesley).
25. G.E.P. Box y G.M. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden Day).
26. R.N. Bracewell, *The Fourier Transform y Its Applications*, 2a. ed., rev. (Nueva York: McGraw-Hill).
27. L. Brillouin, *Science and Information Theory*, 2a. ed. (Nueva York: Academic Press).
28. K.W. Cattermole, *Principles of Pulse-code Modulation* (Nueva York: American Elsevier).

29. W.Y. Chen, *DSL: Simulation Techniques and Standards Development for Digital Subscriber Line Systems* (Indianapolis, Ind.: Macmillan Technical Publishing).
30. J.M. Cioffi, *Digital Data Transmission*, EE379C Course Textbook, Stanford University.
31. G.C. Clark, Jr., y J.B. Cain, *Error-correction Coding for Digital Communications* (Nueva York: Plenum Publishers).
32. L.W. Couch, *Digital and Analog Communication Systems*, 5ª. ed. (Englewood Cliffs, N.J.: Prentice-Hall).
33. T.M. Cover y J.A. Thomas, *Elements of Information Theory* (Nueva York: Wiley).
34. H. Cramér y M.R. Leadbetter, *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications* (Nueva York: Wiley).
35. W.B. Davenport, Jr., y W.I. Root, *An Introduction to the Theory of Random Signals and Noise* (Nueva York: McGraw-Hill).
36. R.C. Dixon, *Spread Spectrum Systems*, 2a. ed. (Nueva York: Wiley).
37. R.C. Dixon (editor), *Spread Spectrum Techniques* (Nueva York, IEEE Press).
38. L.J. Doob, *Stochastic Processes* (Nueva York: Wiley).
39. J.J. Downing, *Modulation Systems and Noise* (Englewood Cliffs, N.J.: Prentice-Hall).
40. W. Feller, *An Introduction to Probability Theory and its Application*, vol. 1, 3a. ed. (Nueva York: Wiley).
41. T.L. Fine, *Theories of Probability: An Examination of Foundations* (Nueva York: Academic Press).
42. L.E. Franks (editor), *Data Communication: Fundamentals of Baseband Transmission* (Dowden, Hutchison y Ross).
43. L.E. Franks, *Signal Theory* (Englewood Cliffs, N.J.: Prentice-Hall).
44. R.L. Freeman, *Telecommunications Transmission Handbook*, 4a. ed. (Nueva York: Wiley).
45. R.M. Gagliardi, *Introduction to Communications Engineering*, 2a. ed. (Nueva York: Wiley).
46. R.G. Gallager, *Information Theory y Reliable Communication* (Nueva York: Wiley).
47. R.G. Gallager, *Low-Density Parity-Check Codes* (Cambridge, Mass.: MIT Press).
48. F.M. Gardner, *Phaselock Techniques*, 2a. ed. (Nueva York: Wiley).
49. V.K. Gary y J.E. Wilkes, *Principles & Applications of GSM* (Englewood Cliffs, N.J.: Prentice-Hall).
50. A. Gersho y R.M. Gray, *Vector Quantization and Signal Compression* (Boston, Mass.: Kluwer Academic).
51. J.D. Gibson (editor), *The Mobile Communications Handbook* (Piscataway, N.J.: IEEE Press).
52. R.D. Gitlin, J.F. Hayes y S.B. Weinstein, *Data Communications Principles* (Nueva York: Plenum).
53. B. Goldberg y H.S. Bennett (editores), *Communication Channels: Characterization and Behavior* (Nueva York: IEEE Press).
54. S.W. Golomb (editor), *Digital Communications with Space Applications* (Englewood Cliffs, N.J.: Prentice-Hall).
55. S.W. Golomb, *Shift Register Sequences* (San Francisco: Holden-Day).
56. R.M. Gray y L.D. Davisson, *Random Processes: A Mathematical Approach for Engineers* (Englewood Cliffs, N.J.: Prentice-Hall).
57. P.E. Green, Jr., *Computer Network Architectures and Protocols* (Nueva York: Plenum).
58. P.E. Green, Jr., *Fiber Optic Networks* (Englewood Cliffs, N.J.: Prentice-Hall).
59. M.S. Gupta (editor), *Electrical Noise: Fundamentals and Sources* (Nueva York: IEEE Press).
60. R.W. Hamming, *Coding and Information Theory* (Englewood Cliffs, N.J.: Prentice-Hall).
61. S. Haykin, *Communication Systems*, 3a. ed. (Nueva York: Wiley).
62. S. Haykin, *Adaptive Filter Theory*, 3a. ed. (Englewood Cliffs, N.J.: Prentice-Hall).
63. S. Haykin y B. Van Veen, *Signals and Systems* (Nueva York: Wiley).
64. C. Heegard y S.B. Wicker, *Turbo Coding* (Boston, Mass.: Kluwer Academic Publishers).
65. G. Held, *The Complete Modem Reference*, 3a. ed. (Nueva York: Wiley).
66. C.W. Helstrom, *Statistical Theory of Signal Detection* (Elmsford, N.Y.: Pergamon Press).
67. C.W. Helstrom, *Probability and Stochastic Processes for Engineers*, 2a. ed. (Nueva York: Macmillan).
68. K. Henney (editor), *Radio Engineering Handbook* (Nueva York: McGraw-Hill).

69. J.K. Holmes, *Coherent Spread Spectrum Systems* (Nueva York: Wiley).
70. W.C. Jakes, Jr. (editor), *Microwave Mobile Communications* (Nueva York: Wiley).
71. N.S. Jayant y P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video* (Englewood Cliffs, N.J.: Prentice-Hall).
72. N.S. Jayant (editor), *Waveform Quantization and Coding* (Nueva York: IEEE Press).
73. H. Jeffreys, Sir, *Theory of Probability*, 3a. ed. (Oxford: Clarendon Press).
74. H. Jeffreys, Sir, y B.S. Jeffreys, *Methods of Mathematical Physics*, 3a. ed. (Cambridge University Press).
75. M.C. Jeruchim, B. Balaban y J. S. Shanmugan, *Simulation of Communication Systems* (Nueva York: Plenum).
76. E.C. Jordan y K.G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2a. ed. (Englewood Cliffs, NJ.: Prentice Hall).
77. A. Khintchin, *Mathematical Foundations of Information Theory* (Nueva York: Dover).
78. A.N. Kolmogorov, *Foundations of the Theory of Probability* (Nueva York: Chelsea Publishing).
79. V.A. Kotel'nikov, *The Theory of Optimum Noise Immunity* (Nueva York: McGraw-Hill).
80. J.D. Kraus, *Antennas* (Nueva York: McGraw-Hill).
81. S. Kullback, *Information Theory and Statistics* (Nueva York: Dover).
82. P. Lafrance, *Fundamental Concepts in Communication* (Englewood Cliffs, N.J.: Prentice-Hall).
83. B.P. Lathi, *Modern Digital and Analog Communication Systems*, 2a. ed. (Oxford University Press).
84. I. Lebow, *Information Highways and Byways* (Piscataway, N.J.: IEEE Press).
85. E.A. Lee y D.G. Messerschmitt, *Digital Communication*, 2a. ed. (Boston, Mass.: Kluwer Academic).
86. J.S. Lee y L.E. Miller, *CDMA Systems Engineering Handbook* (Boston, Mass.: Artech House Publishers).
87. Y.W. Lee, *Statistical The of Communication* (Nueva York: Wiley).
88. W.C.(Y.) Lee, *Mobile Communications Engineering* (Nueva York: McGraw-Hill).
89. A. León-García, *Probability and Random Processes for Electrical Engineering*, 2a. ed. (Reading, Mass.: Addison-Wesley).
90. C.R. Lewart, *The Ultimate Modem Handbook* (Englewood Cliffs, N.J.: Prentice-Hall).
91. S. Lin y D.J. Costello, Jr., *Error Control Coding. Fundamentals and Applications* (Englewood Cliffs, N.J.: Prentice-Hall).
92. W.C. Lindsey, *Synchronization Systems in Communication and Control* (Englewood Cliffs, N.J.: Prentice-Hall).
93. W.C. Lindsey y M.K. Simon (editores), *Phase-locked Loops and Their Applications* (Nueva York: IEEE Press).
94. W.C. Lindsey y M.K. Simon, *Telecommunication Systems Engineering* (Englewood Cliffs, N.J.: Prentice-Hall).
95. M. Loève, *Probability Theory* (Princeton, N.J.: Van Nostrand).
96. R.W. Lucky, *Silicon Dreams: Information, Man, and Machine* (Nueva York: St. Martin's Press).
97. R.W. Lucky, J. Salz y E.J. Weldon, Jr., *Principles of Data Communication* (Nueva York: McGrawHill).
98. F.J. MacWilliams y N.J.A. Sloane, *The Theory of Error-correcting Codes* (Amsterdam: North-Holland).
99. V.K. Madisetti y D.B. Williams (editores), *The Digital Signal Processing Handbook* (Piscataway, N.J.: IEEE Press).
100. R.J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory* (Nueva York/Berlín: Springer-Verlag).
101. J.C. McDonald (editor), *Fundamentals of Digital Switching*, 2a. ed. (Nueva York: Plenum).
102. R. McDonough y A.D. Whalen, *Detection of Signals in Noise*, 2a. ed. (Nueva York: Academic Press).
103. R.J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication* (Reading, Mass.: Addison-Wesley).
104. A. Mengali y N. D'Andrea, *Synchronization Techniques for Digital Receivers* (Nueva York: Plenum).
105. D.J.G. Mestdagh, *Fundamentals of Multiaccess Optical Fiber Networks* (Boston, Mass.: Artech House Publishers).

106. C.H. Meyer y S.M. Matyas, *Cryptography: A New Dimension in Computer Data Security* (Nueva York: Wiley).
107. H. Meyr y G. Ascheid, *Synchronization in Digital Communications*, vol. 1 (Nueva York: Wiley).
108. H. Meyr, M. Moeneclaey y S.A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing* (Nueva York: Wiley).
109. A.M. Michelson y A.H. Levesque, *Error-control Techniques for Digital Communication* (Nueva York: Wiley).
110. D. Middleton, *An Introduction to Statistical Communication Theory* (Nueva York: McGraw-Hill).
111. J.G. Nellist, *Understanding Telecommunications and Lightwave Systems: An Entry Level Guide* (Piscataway, N.J.: IEEE Press).
112. C.F.J. Overhage (editor), *The Age of Electronics* (Nueva York: McGraw-Hill).
113. P.F. Panter, *Modulation, Noise and Spectral Analysis, Applied to Information Transmission* (Nueva York: McGraw-Hill).
114. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2a. ed. (Nueva York: McGraw-Hill).
115. J.D. Parsons, *The Mobile Radio Propagation Channel* (Nueva York: Wiley).
116. K. Pahlavan y A.H. Levesque, *Wireless Information Networks* (Nueva York: Wiley).
117. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Mateo, Calif: Morgan Kaufman Publishers).
118. W.W. Peterson y E.J. Weldon, Jr., *Error Correcting Codes*, 2a. ed. (Cambridge, Mass.: MIT Press).
119. J.R. Pierce y A.M. Noll, *Signals: The Science of Telecommunications* (Nueva York: Scientific American Library).
120. J.R. Pierce, *Symbols, Signals and Noise: The Nature and Process of Communication* (Nueva York: Harper).
121. H.V. Poor, *An Introduction to Signal Detection and Estimation*, 2a. ed. (Nueva York/Berlín: Springer-Verlag).
122. T. Pratt y C.W. Bostian, *Satellite Communications* (Nueva York: Wiley).
123. J.G. Proakis, *Digital Communications*, 3a. ed. (Nueva York: McGraw-Hill).
124. L.R. Rabiner y R.W. Schafer, *Digital Processing of Speech Signals* (Englewood Cliffs, N.J.: Prentice-Hall).
125. K.R. Rao y P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications* (Nueva York: Academic Press).
126. T.S. Rappaport, *Smart Antennas* (Piscataway, N.J.: IEEE Press).
127. T.S. Rappaport, *Wireless Communications: Principles and Practice* (Piscataway, N.J.: IEEE Press).
128. S.O. Rice, "Noise in FM receivers", en M. Rosenblatt (editor), *Proceedings of the Symposium on Time Series Analysis* (Nueva York: Wiley), pp. 395-411.
129. J.H. Roberts, *Angle Modulation: The Theory of Systems Assessment*, IEE Communication Series 5 (Londres: Institution of Electrical Engineers).
130. H.E. Rowe, *Signals and Noise in Communication Systems* (Princeton, N.J.: Van Nostrand).
131. D.J. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information* (Nueva York: Wiley).
132. C. Schlegel, *Trellis Coding* (Piscataway, N.J.: IEEE Press).
133. M. Schwartz, W.R. Bennett y S. Stein, *Communication Systems and Techniques* (Nueva York: McGraw-Hill).
134. M. Schwartz, *Information Transmission, Modulation and Noise: A Unified Approach*, 3a. ed. (Nueva York: McGraw-Hill).
135. M. Schwartz, *Telecommunication Networks: Protocols, Modeling, and Analysis* (Reading, Mass.: Addison Wesley).
136. K.S. Shanmugan, *Digital and Analog Communication Systems* (Nueva York: Wiley).

137. C.E. Shannon y W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press).
138. G.J. Simmons (editor), *Contemporary Cryptology: The Science of Information Integrity* (Piscataway, N.J.: IEEE Press).
139. M.K. Simon, J.K. Omura, R.A. Scholtz y B.K. Levitt, *Spread Spectrum Communications*, vols. I, II, y III (Nueva York: Computer Science Press).
140. B. Sklar, *Digital Communications: Fundamentals and Applications* (Englewood Cliffs, N.J.: Prentice-Hall).
141. D. Slepian (editor), *Key Papers in the Development of Information Theory* (Nueva York: IEEE Press).
142. N.J.A. Sloane y A.D. Wyner, *Claude Shannon: Collected Papers* (Piscataway, N.J.: IEEE Press).
143. D.R. Smith, *Digital Transmission Systems* (Princeton, N.J.: Van Nostrand Reinhold).
144. I.S. Sokolnikoff y R.M. Redheffer, *Mathematics of Physics and Modern Engineering* (Nueva York: McGraw-Hill).
145. J.J. Spilker, Jr., *Digital Communications by Satellite* (Englewood Cliffs, N.J.: Prentice-Hall).
146. W. Stallings, *ISDN and Broadband ISDN*, 2a. ed. (Nueva York: Macmillan).
147. T. Starr, J.M. Cioffi y P.J. Silverman, *Understanding Digital Subscriber Line Technology* (Englewood Cliffs, N.J.: Prentice-Hall).
148. R. Steele, *Delta Modulation Systems* (Nueva York: Wiley).
149. R. Steele y L. Hanzo (editores), *Mobile Radio Communications*, 2a. ed. (Nueva York: Wiley).
150. J.J. Stiffler, *Theory of Synchronous Communications* (Englewood Cliffs, N.J.: Prentice-Hall).
151. E.D. Sunde, *Communications Systems Engineering Theory* (Nueva York: Wiley).
152. A.S. Tanenbaum, *Computer Networks*, 2a. ed. (Englewood Cliffs, N.J.: Prentice-Hall).
153. S. Tantaratana y K.M. Ahmed, *Wireless Applications of Spread Spectrum Systems: Selected Readings* (Piscataway, N.J.: IEEE Press).
154. T.M. Thompson, *From Error-correcting Codes Through Sphere Packings to Simple Groups* (The Mathematical Association of America, Washington D.C.).
155. D.J. Torrieri, *Principles of Military Communication Systems*, 2a. ed. (Boston, Mass.: Artech House Publishers).
156. G.L. Turin, *Notes on Digital Communications* (Princeton, N.J.: Van Nostrand-Reinhold).
157. A. Van der Ziel. *Noise: Source, Characterization, Measurement* (Englewood Cliffs, N.J.: Prentice-Hall).
158. H.F. Vanlandingham, *Introduction to Digital Control Systems* (Nueva York: Macmillan).
159. H.C.A. van Tilborg, *An Introduction to Cryptology* (Boston, Mass.: Kluwer).
160. H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, Parte I (Nueva York: Wiley).
161. A.J. Viterbi, *Principles of Coherent Communication* (Nueva York: McGraw-Hill).
162. A.J. Viterbi y J.K. Omura, *Principles of Digital Communication and Coding* (Nueva York: McGraw-Hill).
163. G.N. Watson, A *Treatise in the Theory of Bessel Functions*, 2a. ed. (Nueva York: Cambridge University Press).
164. N. Wax (editor), *Selected Papers on Noise and Stochastic Processes* (Nueva York: Dover Publications).
165. E.T. Whittaker y G.N. Watson, *A Course in Modern Analysis*, 4a. ed. (Nueva York: Cambridge University Press).
166. S.B. Wicker y V.K. Bhargava (editores), *Reed-Solomon Codes* (Piscataway, N.J.: IEEE Press).
167. B. Widrow y S.D. Stearns, *Adaptive Signal Processing* (Englewood Cliffs, N.J.: Prentice-Hall).
168. N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications* (Nueva York: Wiley).
169. S.G. Wilson, *Digital Modulation and Coding* (Englewood Cliffs, N.J.: Prentice-Hall).
170. E. Wong, *Stochastic Processes in Information and Dynamical Systems* (Nueva York: McGraw-Hill).
171. P.M. Woodward, *Probability and Informattion Theory, with Applications to Radar*, 2a. ed. (Elmsford, N.Y.: Pergamon Press).

172. J.M. Wozencraft e I.M. Jacobs, *Principles of Communication Engineering* (Nueva York: Wiley).
173. W.W. Wu, *Elements of Digital Satellite Communication*, vol. I (Nueva York: Computer Science Press).
174. C.R. Wylie y L.C. Barrett, *Advanced Engineering Mathematics*, 5th ed. (Nueva York: McGraw-Hill).
175. R.D. Yates y D.J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers* (Nueva York: Wiley).
176. J.H. Yuen (editor), *Deep Space Telecommunications Systems Engineering* (Nueva York: Plenum).
177. R.E. Ziemer y R.L. Peterson, *Digital Communications and Spread Spectrum Systems* (Nueva York: Macmillan).
178. R.E. Ziemer y W.H. Tranter, *Principles of Communications*, 3a. ed. (Boston, Mass.: Houghton Miflin).

ARTÍCULOS, REPORTES, PATENTES[1]

179. M.R. Aaron y D.W. Tufts, "Intersymbol interference and error probabdity", *IEEE Trans. on InformationTheory* vol. IT-12, pp. 26-34.
180. J.E. Abate, "Linear and adaptive delta modulation", *Proceedings of the IEEE*, vol. 55, pp. 298-308.
181. A.N. Akansu, P. Duhamel, X. Lin y M. de Courville, "Orthogonal transmultiplexers in communications: A review", *IEEE Transactions on Signal Processing*, vol. 46, pp. 979-995.
182. Y. Akaiwa y Y. Nagata, "Highly efficient digital mobile communications with a linear modulation method", *IEEE Journal on Selected Areas in Communications*, vol. SAC-5, pp. 890-895.
183. O. Al-Shaykh, R. Neff, D. Taubman y A. Zakhour, "Video sequence compression." En V.K. Madisetti y D.B. Williams (editores), *The Digital Signal Processing Handbook*, CRC Press, pp. 55-1-55-19.
184. F. Amoroso, "The bandwidth of digital data signals", *IEEE Communications Magazine*, vol. 18, no. 6, pp. 13-24.
185. J.B. Anderson y D.P. Taylor, "A bandwidth-efficient class of signal space codes", *IEEE Transactions on Information Theory*, vol. IT-24, pp. 703-712.
186. R.R. Anderson y J. Salz, "Spectra of digital FM", *Bell System Tech. J.*, vol. 44, pp. 1165-1189.
187. R. Arens, "Complex processes for envelopes of normal noise", *IRE Trans. on Information Theory*, vol. IT-3, pp. 204-207.
188. E.H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation", *Proceedings of the IRE*, vol. 24, pp. 689-740.
189. E. Arthurs y H. Dym, "On the optimum detection of digital signals in the presence of white Gaussian noise-A geometric interpretation and a study of three basic data transmission systems", *IRE Trans. on Communication Systems*, vol. CS-10, pp. 336-372.
190. B.S. Atal y J.R. Remde, "A New model of LPC excitation for producing natural-sounding speech at low bit rates", *Proc. ICASSP, '82*, pp. 614-17.
191. B.S. Atal y M.R. Schroeder, "Stochastic coding of speech signals at very low bit rates", IEEE International Conference on Communications.
192. M. Austin, "Decision-feedback equalization for digital communication over dispersive channels", *MIT Research Laboratory of Electronics Technical Report* 461.
193. E. Ayanoglu, N.R. Dagdeviren, J.E. Mazo y R. Saltzberg, "High-speed modem synchronized to remote codec", patente estadounidense 5-394-437.
194. E. Ayanoglu, N.R. Dagdeviren, G.D. Golden y J.E. Mazo, "An equalizer design technique for the PCM modem: a new modem for the digital public switched network", *IEEE Transactions on Communications*, vol. 46, pp. 763-774.
195. L.R. Bahl, J. Cocke, F. Jelinek y J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate", *IEEE Transactions on Information Theory*, vol. IT-20, pp. 284-287.
196. G. Battail, "Coding for the Gaussian channel: the promise of weighted output decoding", *International J. Satellite Communications*, vol. 7, pp. 183-192.

197. G. Battail, "Pondération des symbols décodés par l'algorithme de Viterbi", *Ann. Télécommunication*, vol. 42, pp. 31-38.
198. E. Bedrosian, "The analytic signal representation of modulated waveforms", *Proceedings of the IRE*, vol. 50, pp. 2071-2076.
199. P.A. Bello, "Characterization of randomly time-variant linear channels", *IEEE Transactions on Communication Systems*, vol. CS-11, pp. 360-393.
200. S. Benedetto y G. Montorsi, "Unveiling turbo codes: Some results en parallel concatenated coding schemes", *IEEE Transactions on Information Theory*, vol. 42, pp. 409-428.
201. W.R. Bennett, "Spectra of quantized signals", *Bell System Tech. J.*, vol. 27, pp. 446-472.
202. N. Benvenuto y otros, "The 32 kb/s ADPCM coding standard", *AT&T Technical Journal*, vol. 65, pp. 12-22.
203. C. Berrou y A. Glavieux, "Near optimum error correcting coding y decoding: turbo codes", *IEEE Transactions on Communications*, vol. 44, pp. 1261-1271.
204. C. Berrou y A. Glavieux, "Reflections on the Prize Paper: Near optimum error-correcting coding and decoding turbo codes", *IEEE Information Theory Society Newsletter*, vol. 48, no. 2, p. 1 y pp. 24-31.
205. C. Berrou, A. Glavieux y P. Thitmajshima, "Near Shannon limit error-correction coding and turbo codes", *International Conference on Communications*, pp. 1064-1090, Ginebra, Suiza.
206. V.K. Bhargava, "Forward error correction schemes for digital communications", *IEEE Communications Magazine*, vol. 21, no. 1, pp. 11-19.
207. R.C. Bose y D.K. Ray-Chaudhuri, "On a class of error correcting binary group codes", *Information and Control*, vol. 3, pp. 68-79.
208. K. Brandenburg y G. Stoll, "ISO-MPEG-1 Audio: A generic standard for coding of high-quality digital audio", *Journal of the Audio Engineering Society*, vol. 42, pp. 780-792.
209. D.G. Brennan, "Linear diversity combining techniques", *Proceedings of the IRE*, vol. 47, pp. 1075-1102.
210. A. Buzo, A.H. Gray, Jr., R.M. Gray y J.D. Markel, "Speech coding based upon vector quantization", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, pp. 562-574.
211. C.R. Cahn, "Combined digital phase and amplitude modulation communication systems", *IRE Transactions on Communication Systems*, vol. CS-8, pp. 150-155.
212. J.R. Carson, "Notes on the theory of modulation", *Proceedings of the IRE*, vol. 10, pp. 57-64.
213. J.R. Carson y T.C. Fry, "Variable frequency electric circuit theory with application to the theory of frequency modulation", *Bell System Tech. J.*, vol. 16, pp. 513-540.
214. E.F. Casas y C. Leung, "OFDM for data communication over mobile radio FM channels", *IEEE Transactions on Communications*, vol. 39, pp. 783-793.
215. J.G. Chaffee, "The application of negative feedback to frequency-modulation systems", *Bell System Tech. J.*, vol. 18, pp. 404-437.
216. R.W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission", *Bell System Tech. J.*, vol. 45, pp. 1775-1796.
217. W.Y. Chen, G.H. Im y J.J. Werner, "Design of digital carrierless AM/PM transceivers", *Standard Project, T1 E1.4/92-149, AT&T and Bellcore.*
218. S. Chennakeshu y G.J. Sauliner, "Differential detection of $\pi/4$-shifted-DQPSK for digital cellular radio", *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 46-57.
219. J.M. Cioffi, V. Oksman, J.-J. Wemer, T. Pollet, P.M.P. Spruyt, J.S. Chow y K.S. Jacobsen, "Veryhigh-speed digital subscriber lines", *IEEE Communications Magazine*, vol. 37, pp. 72-79.
220. L.J. Cimini, Jr., y Y. Li, "Orthogonal frequency division multiplexing for wireless communications", notas tutoriales, TU18, Conferencia Internacional sobre Comunicaciones '99, Vancouver, British Columbia, Canada.
221. A.C. Clarke, "Extraterrestrial relays", *Wireless World*, vol. 51, pp. 305-308.

222. C.E. Cook y H.S. Marsh, "An introduction to spread spectrum", *IEEE Communications Magazine*, vol. 21, no. 2, pp. 8-16.
223. J.P. Costas, "Synchronous communications", *Proceedings of the IRE*, vol. 44, pp. 1713-1718.
224. J.P. Costas, "Poisson, Shannon, and the radio amateur", *Proceedings of the IRE*, vol. 47, pp. 2058-2068.
225. M.G. Crosby, "Frequency modulation noise characteristics", *Proceedings of the IRE*, vol. 25, pp. 472-514.
226. C.C. Cutler, "Differential quantization of communication signals", patente estadounidense 2-505-361.
227. C.L. Dammann, L.D. McDaniel y C.L. Maddox, "D2 channel bank-Multiplexing and coding", *Bell System Tech. J.* vol. 51, pp. 1675-1699.
228. F. Daneshgaran y M. Mondin, "Design of interleavers for turbo codes: Iterative interleaver growth algorithms of polynomial complexity", *IEEE Transactions on Information Theory*, vol. 45, pp. 1845-1859.
229. R. deBuda, "Coherent demodulation of frequency-shift keying with low deviation ratio", *IEEE Trans. on Communications*, vol. COM-20, pp. 429-435.
230. F.E. DeJager, "Deltamodulation, a method of PCM transmission using the 1-unit code", *Phillips Research Reports*, vol. 7, pp. 442-46.
231. F.E. DeJager y C.B. Dekker, "Tamed frequency modulation: A novel method to achieve spectrum economy in digital transmission", *IEEE Transactions on Communications*, vol. COM-26, pp, 534-542.
232. J.A. Develet, "A threshold criterion for phase-lock demodulation", *Proceedings of the IEEE*, vol. 51, pp. 349-356.
233. W. Diffie y M.E. Hellman, "New directions in cryptography", *IEEE Transactions on Information Theory*, vol. IT-22, pp. 644-654.
234. W. Diffie y M.E. Hellman, "Privacy and authentication: An introduction to cryptography", *Proceedings of the IEEE*, vol. 67, pp. 397-427.
235. D. Divsalar, "Turbo codes", MILCOM 96 tutorial, San Diego.
236. M.1. Doelz y E.H. Heald, "Minimum shift data communication system", patente estadounidense 2977417.
237. R.M. Dolby, "An audio reduction system", *Journal of the Audio Engineering Society*, vol. 15, p. 383.
238. J. Dungundji, "Envelopes and pre-envelopes of real wave-forms", *IRE Transactions on Information Theory*, vol. IT-4, pp. 53-57.
239. P. Elias, "Coding for noisy channels", *IRE Convention Record*, Part 4, pp. 37-46.
240. L.H. Enloe, "Decreasing the threshold in FM by frequency feedback", *Proceedings of the IRE*, vol. 50, pp. 18-30.
241. V.M. Eyuboglu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving", *IEEE Transactions on Communications*, vol. COM-36, pp. 401-409.
242. D.D. Falconer, "Carrierless AM/PM", *Bell Laboratories, Internal Memorandum*.
243. K. Feher, "MODEMS for emerging digital cellular-mobile radio system", *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 355-365.
244. J.L. Flanagan, M.R. Schroeder, B.S. Atal, R.E. Crochiere, N.S. Jayant y J.M. Tribolet, "Speech coding", *IEEE Transactions on Communications*, vol. COM-27, pp. 710-737.
245. B. LeFloch, R. Halbert-Lassalle y D. Castelain, "Digital sound broadcasting to mobile receivers", *IEEE Transactions of Broadcasting*, vol. 35, pp. 493-503.
246. G.D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference", *IEEE Transactions on Information Theory*, vol. IT-18, pp. 363-378.
247. G.D. Forney, Jr. "The Viterbi algorithm", *Proceedings of the IEEE*, vol. 61, pp. 268-278.
248. G.D. Forney, Jr., y M.V. Eyuboglu, "Combined equalization and coding using precoding", *IEEE Communications Magazine*, vol. 29, no. 12, pp. 25-34.

249. G.D. Forney, Jr., L. Brown, M.V. Eyuboglu y J.L. Moran III, "The V.34 high-speed modem standard", *IEEE Communications Magazine*, pp. 28-93.
250. L.E. Franks, "Carrier and bit synchronization in data communications-A tutorial review", *IEEE Transactions on Communications*, vol. COM-28, pp. 1107-1121.
251. B.J. Frey y D.J.C. MacKay, "Irregular turbocodes", *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, Illinois.
252. H.T. Friis, "Noise figures in radio receivers", *Proceedings of the IRE*, vol. 32, pp. 419-422.
253. K.E. Fulz y D.B. Penick, "Tl carrier system", *Bell System Tech. J.*, vol. 44, pp. 1405-1451.
254. D. Gabor, "Theory of communications", *Journal of IEE* (Londres), vol. 93, Parte III, pp. 429-457.
255. D.L. Gall, "MPEG: a video compression standard for multimedia applications", *Communications of the ACM*, vol. 34, pp. 47-58.
256. R.G. Gallager, "Low-density parity-check codes", *IRE Transactions on Information Theory*, vol. 8, pp. 21-28.
257. W.A. Gardner, "Introduction to Einstein's contribution to time-series analysis", *IEEE ASSP Magazine*, vol. 4, pp. 4-5.
258. W.A. Gardner y L.E. Franks, "Characterization of cyclostationary random signal processes", *IEEE Transactions on Information Theory*, vol. IT-21, pp. 4-14.
259. D.A. George, "Matched filters for interfering signals", *IEEE Transactions on Information Theory*, vol. IT-11, pp. 153-154.
260. A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission", *Bell System Tech. J.*, vol. 48, pp. 55-70.
261. R.A. Gibby y J.W. Smith, "Some extensions of Nyquist's telegraph transmission theory", *Bell Systems Tech. J.*, vol. 44, pp. 1487-1510.
262. R.D. Gitlin y E.Y. Ho, "The performance of staggered quadrature amplitude modulation in the presence of phase jitter", *IEEE Transactions on Communications*, vol. COM-23, pp. 348-352.
263. M.J.E. Golay, "Note on digital coding", *Proceedings of the* IRE, vol. 37, p. 657.
264. M.J.E. Golay, "Binary coding", *IRE Transactions on Information Theory*, vol. PGIT-4, pp. 23-28.
265. R. Gold, "Optimal binary sequences for spread spectrum multiplexing", *IEEE Transactions on Information Theory*, vol. IT-13, pp. 619-621.
266. R. Gold, "Maximal recursive sequences with 3-valued recursive cross correlation functions", *IEEE Transactions on Information Theory*, vol. IT-14, pp. 154-156.
267. B. Goode, "Scanning the issue: Special issue on global information infrastructure", *Proceedings of the IEEE*, vol. 85, pp. 1883-1886.
268. R.M. Gray, "Vector quantization", *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4-29.
269. W.J. Gruen, "Theory of AFC synchronization", *Proceedings of the IRE*, vol. 41, pp. 1043-1048.
270. P. Guinand y J. Lodge, "Trellis termination for turbo encoders", *Proceedings of 18th Biennial Symposium on Communications, Queen's University*, Kingston, Canadá.
271. D.W. Hagelbarger, "Recurrent codes: Easily mechanized, burst-correcting binary codes", *Bell System Tech.* J., vol. 38, pp. 969-984.
272. J. Hagenauer, E. Offer y L. Papke, "Iterative decoding of binary block and convolutional codes", *IEEE Transactions on Information Theory*, vol. 42, pp. 429-445.
273. J. Hagenaucr y P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", *IEEE Globecom* 89, pp. 47.11-47.17, Dallas, Texas.
274. R.W. Hamming, "Error detecting and error correcting codes", *Bell System Tech. J.*, vol. 29, pp. 147-160.
275. J.C. Hancock y R.W. Lucky, "Performance of combined amplitude and phase-modulated communication systems", IRE *Transactions on Communication Systems*, vol. CS-8, pp. 232-237.
276. H.H. Hanning y J.W. Pan, "D2 channel bank system aspects", *Bell System Tech. J.*, vol. 51, pp. 1641-1657.

277. H. Harashima y H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference", *IEEE Transactions on Communications*, vol. COM-20, pp. 774-779.
278. T.V.L. Hartley, "Transmission of information", *Bell System Tech. J.*, vol. 7, pp. 535-563.
279. F.S. Hill, Jr., "On time-domain representations for vestigial sideband signals", *Proceedings of the IEEE*, vol. 62, pp. 1032-1033.
280. D.A. Huffman, "A method for the construction of minimum redundancy codes", *Proceedings of the IRE*, vol. 40, pp. 1098-1101.
281. P.A. Humblet y M.G. Troulis, "The information driveway", *IEEE Communications Magazine*, pp. 64-68.
282. G.-H. Im y J.-J. Werner, "Bandwidth-efficient digital transmission over unshielded twisted-pair wiring", *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1643-1655.
283. H. Insoe, Y. Yasuda y J. Murakami, "A telemetering system by code modulation: Δ-Σ modulation", *IRE Transactions on Space Electronics and Telemetry*, vol. SET-8, pp. 204-209.
284. M. Ishizuka y K. Hirade, "Optimum Gaussian filter and deviated-frequency locking scheme for coherent detection of MSK", *IEEE Transactions on Communications*, vol. COM-28, pp. 850-857.
285 I.M. Jacobs, "Practical applications of coding", *IEEE Transactions on Information Theory*, vol. IT-20, pp. 305-310.
286. N.S. Jayant, "Adaptive delta modulation with a one-bit memory", *Bell System Tech. J.*, vol. 49, pp. 321-342.
287. N.S. Jayant, "Digital coding of speech waveforms, PCM, DPCM and DM quantizers", *Proceedings of the IEEE*, vol. 62, pp. 611-632.
288. N.S. Jayant, "Coding speech at low bit rates", *IEEE Spectrum*, vol. 23, no. 8, pp. 58-63.
289. A.J. Jerri, "The Shannon sampling theorem-its various extensions and applications: A tutorial review", *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565-1596.
290. J.B. Johnson, "Thermal agitation of electricity in conductors", *Physical Review*, segunda serie, vol. 32, pp. 97-109.
291. P. Kabal y S. Pasupathy, "Partial-response signaling", *IEEE Transactions on Communications*, vol. COM-23, pp. 921-934.
292. I. Kalet, "The multitone channel", *IEEE Transactions on Communications*, vol. 37, pp. 119-124.
293. I. Kalet, J.E. Mazo y B.R. Saltzberg, "The capacity of PCM voiceband channels", *IEEE International Conference on Communications*, pp. 507-511, Ginebra, Suiza.
294. H. Kaneko, "A unified formulation of segment companding laws and synthesis of codes and digital companders", *Bell System Tech. J.*, vol. 49, pp. 1555-1588.
295. A.I. Khintchine, "Korrelationstheorie der stationören stochastischen Prozese", *Mathematiche Annalen*, vol. 1, 109, pp. 415-458.
296. H. Kobayashi, "Correlative level coding and maximum-likelihood decoding", *IEEE Transactions on Information Theory*, vol. IT-17, pp. 586-594.
297. R. Kohno, "Spatial and temporal communication theory using adaptive antenna array", *IEEE Personal Communications*, pp. 28-35.
298. E.T. Kretzmer, "Generalization of a technique for binary data communications", *IEEE Transactions on Communication Technology*, vol. COM-14, pp. 67-68.
299. F.R. Kschischang y B.J. Frey, "Interactive decoding of compound codes by probability propagation in graphical models", *IEEE Journal on Selected Areas in Communication*, vol. 16, pp. 219-230.
300. J.W. Lechleider, "Line codes for digital subscriber lines", *IEEE Communications Magazine*, vol. 27, pp. 25-32.
301. B.M. Leiner, V.G. Cerf, D.D. Clark, R.E. Kohn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts y S. Wolff, "A brief history of the Internet", *Commun. ACM*, vol. 40, pp. 102-108.
302. A. Lender, "The duobinary technique for high-speed data transmission", *IEEE Transactions on Communications and Electronics*, vol. 82, pp. 214-218.

303. A. Lender, "Correlative digital communication techniques", *IEEE Transactions on Communication Technology*, vol. COM-12, pp. 128-135.
304. A. Lender, "Correlative level coding for binary-data transmission", *IEEE Spectrum*, vol. 3, no. 2, pp.104-115.
305. N.-S. Lin y C.-P.J. Tzeng, "Full-duplex data over local loops", *IEEE Communications Magazine*, vol. 26, pp. 31-42.
306. S. Lin, D.J. Costello y M.J. Miller, "Automatic-repeat-request error control schemes", *IEEE Communications Magazine*, vol. 22, no. 12, pp. 5-16.
307. Y. Linde, A. Buzo y R.M. Gray, "An algorithm for vector quantizer design", *IEEE Trans. on Communications*, vol. COM-28, pp. 84-95.
308. D. Linden, "A discussion of sampling theorems", *Proceedings of the IRE*, vol. 47, pp. 1219-1226.
309. C.L. Liu y K. Feher, "Noncoherent derection of $\pi/4$-shifted systems in a CCI-AWGN combined interference environment", *Proceedings of the IEEE 40th Vehicular Technology Conference*, San Francisco.
310. S.P. Lloyd, "Least squares quantization in PCM", nota sin publicar de Bell Laboratories. Este reporte se reprodujo en *IEEE Transactions on Information Theory*, vol. IT-28, pp.129-137.
311. J. Lodge, R.Young, P. Hocher y J. Hagenauer, "Separable MAP 'filters' for the decoding of product and concatenated codes", *Proceedings of the IEEE International Conference on Communications*, pp. 1740-1745, Ginebra, Suiza.
312. R. W. Lucky, "Automatic equalization for digital conmunication", *Bell System Tech. J.*, vol. 44, pp. 547-588.
313. R. W. Lucky, "Techniques for adaptive equalization of digital communication systems", *Bell System Tech. J.*, vol. 45, pp. 255-286.
314. R. Lugannani, "Intersymbol interference and probability of error in digital systems", *IEEE Transactions on Information Theory*, vol. IT-15, pp. 682-688.
315. V.H. MacDonald, "Advanced mobile phone service: the cellular concept", *Bell System Tech. J.*, vol. 58, pp. 15-41.
316. D.J.C. MacKay, "Good error-correcting codes based on very sparse matrices", *IEEE Transactions on Information Theory*, vol. 45, pp. 399-431.
317. D.J.C. MacKay y R.M. Neal, "Near Shannon limit performance of low density parity check codes", *Electronics Letters*, vol. 33, No. 6, pp. 457-458; y vol. 32, no. 18, pp. 1645-1646.
318. D.J.C. MacKay, S.T. Wilson y M.C. Davey, "Comparison of constructions of irregular Gallager codes", *IEEE Transactions on Communications*, vol. 47, pp. 1449-1454.
319. J. Max, "Quantizing for minimum distortion", *IRE Transactions on Information Theory*, vol. IT-6, pp. 7-12.
320. K. Maxwell, "Asymmetric digital subscriber line: Interim technology for the next forty years", *IEEE Communications Magazine*, vol. 34, pp. 100-106.
321. R.J. McEliece, D.J.C. MacKay y J.-F. Cheng, "Turbo coding as an instance of Pearl's belief propagation algorithm", *IEEE Journal on Selected Areas of Communication*, vol. 16, pp. 140-152.
322. D. Mennie, "AM stereo: Five competing options", *IEEE Spectrum*, vol. 15, no. 6, pp. 24-31.
323. M.L. Moher, "Cross-entropy and iterative detection", tesis doctoral, Department of Systems and Computer Engineering, Carleton University, Ottawa, Canadá.
324. M.L. Moher y T.A. Gulliver, "Cross-entropy and iterative decoding", *IEEE Transactions on Information Theory*, vol. 44, pp. 3097-3104.
325. P. Monsen, "Feedback equalization for fading dispersive channels", *IEEE Transactions on Information Theory*, vol. IT-17, pp. 56-64.
326. K.H. Mueller y J.J. Werner, "A hardware efficient passband equalizer structure for data transmissions", *IEEE Transactions on Communications*, vol. COM-30, pp. 538-541.

327. K. Murota y K. Hirade, "GMSK modulation for digital mobile radio telephone", *IEEE Transactions on Communications*, vol. COM-29, pp. 1044-1050.
328. E. Murphy, "Whatever happened to AM stereo?" *IEEE Spectrum*, vol. 25, p. 17.
329. P. Noll, "MPEG digital audio coding standards." En V.K. Madisetti y D.B. Williams (editores), *The Digital Signal Processing Handbook*, Piscataway, N.J.: IEEE Press, pp. 40-1-40-28.
330. D.O. North, "An analysis of the factors which determine signal/noise discrimination in pulsed carrier systems", *Proceedings of the IEEE*, vol. 51, pp. 1016-1027; este artículo es una reproducción de un informe confidencial de RCA que se publicó en 1943.
331. H. Nyquist, "Certain factors affecting telegraph speed", *Bell System Tech. J.*, vol. 3, pp. 324-346.
332. H. Nyquist, "Thermal agitation of electric charge in conductors", *Physical Review*, segunda serie, vol. 32, pp. 110-113.
333. H. Nyquist, "Certain topics in telegraph transmission theory", *Transactions of the AIEE*, vol. 47, pp. 617-644.
334. M.W. Oliphant, "The mobile phone meets the Internet", *IEEE Spectrum*, vol. 36, pp. 20-28.
335. B.M. Oliver, J.R. Pierce y C.E. Shannon, "The philosophy of PCM", *Proceedings of the IRE*, vol. 36, pp. 1324-1331.
336. D.Y. Pan, "Digital audio compression", *Digital Technical Journal*, vol. 5, pp. 1-14.
337. S. Pasupathy, "Nyquist's third criterion", *Proceedings of the IEEE*, vol. 62, pp. 860-861.
338. S. Pasupathy, "Correlative coding-A bandwidth-efficient signaling scheme", *IEEE Communications Magazine*, vol. 15, no. 4, pp. 4-11.
339. S. Pasupathy, "Minimum shift keying-A spectrally efficient modulation", *IEEE Communications Magazine*, vol. 17, no. 4, pp. 14-22.
340. A.J. Paulraj y B.C. Ng, "Space-time modems for wireless personal communications", *IEEE Personal Communications*, pp. 36-48.
341. A.J. Paulraj y C.B. Papadias, "Space-time processing for wireless communications", *IEEE Signal Processing Magazine*, pp. 49-83.
342. R.L. Pickholtz, D.L. Schilling y L.B. Milstein, "Theory of spread-spectrum communications-A tutorial", *IEEE Transactions on Communications*, vol. COM-30, pp. 855-884.
343. R. Price, "Nonlinearly feedback-equalized PAM vs. capacity for noisy filter channels", Conferencia Internacional sobre Comunicaciones, '72, pp. 22.12-22.17, Filadelfia.
344. R. Price y P.E. Green, Jr., "A communication technique for multipath channels", *Proceedings of the IRE*, vol. 46, pp. 555-570.
345. J.G. Proakis, "Advances in equalization for intersymbol interference", *Advances in Communications Systems*, editado por A.J. Viterbi, vol. 4, pp. 123-198, Academic Press.
346. S. Qureshi, "Adaptive equalization", *IEEE Communications Magazine*, vol. 20, no. 2, pp. 9-16.
347. S. Qureshi, "Adaptive equalization", *Proceedings of the IEEE*, vol. 73, pp. 1349-1387.
348. T.A. Ramstad, "Still image compression." En V.K. Madisetti y D.B. Williams (editores), *The Digital Signal Processing Handbook*, Piscataway, N.J.: IEEE Press, pp. 52-1-52-27.
349. I.S. Reed y G. Solomon, "Polynomial codes over certain finite fields", *Journal of SIAM*, vol. 8, pp. 300-304.
350. A.H. Reeves, "The past, present and future of PCM", *IEEE Spectrum*, vol. 12, no. 5, pp. 58-63.
351. S.A. Rhodes, "Effect of noisy phase reference on coherent detection of offset-QPSK signals", *IEEE Transactions on Communications*, vol. COM-22, pp. 1046-1055.
352. S.O. Rice, "Mathematical analysis of random noise", *Bell System Tech. J.*, vol. 23, pp. 282-332; vol. 24, pp. 46-156.
353. S.O. Rice, "Statistical properties of a sine-wave plus random noise", *Bell System Tech. J.*, vol. 27, pp.109-157.
354. S.O. Rice, "Envelopes of narrow-band signals", *Proceedings of the IEEE*, vol. 70, pp. 692-699.

355. T. Richardson, A. Shokrollahi y R. Urbanke, "Design of provably good low-density parity check codes", propuesto en 1999 para *IEEE Transactions on Information Theory.*
356. R.L. Rivest, A. Shamir y L. Adleman, "A method for obtaining digital signatures and public key cryptosystems", *Communications of the* ACM, vol. 21, pp. 120-126.
357. W.L. Root, "Remarks, mostly historical, on signal detection and signal parameter estimation", *Proceedings of the IEEE,* vol. 75, pp. 1446-1457.
358. A. Ruiz, J.M. Cioffi y S. Kasturia, "Discrete multiple tone modulation with coset coding for the spectrally shaped channel", *IEEE Transactions on Communications,* vol. 40, pp. 1012-1029.
359. W.D. Rummler, "A New selective fading model-Application to propagation data", *Bell System Tech. J.,* vol. 58, pp. 1037-1071.
360. B.R. Saltzberg, "Comparison of single-carrier and multitone digital modulation for ADSL applications", *IEEE Communications Magazine,* vol. 36, pp. 114-121.
361. B.R. Saltzberg, "Performance of an efficient parallel data transmission system", *IEEE Transactions on Communication Technology,* vol. COM-15, pp. 805-811.
362. S.D. Sandberg y M.A. Tzannes, "Overlapped discrete multitone modulation for high speed copper wire communications", *IEEE Journal on Selected Areas in Communications,* vol. 13, pp. 1571-1585.
363. D.V. Sarwate y M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences", *Proceedings of the IEEE,* vol. 68, pp. 593-619.
364. B. Sayar y S. Pasupathy, "Nyquist 3 pulse shaping in continuous phase modulation", *IEEE Transactions on Communications,* vol. COM-35, pp. 57-67.
365. H.R. Schindler, "Delta modulation", *IEEE Spectrum,* vol. 7, no. 10, pp. 69-78.
366. R.A. Scholz, "The origins of spread-spectrum communications", *IEEE Transactions on Communications,* vol. COM-30, pp. 822-854.
367. R.A. Scholz, "Notes on spread-spectrum history", *IEEE Transactions on Communications,* vol. COM-31, pp. 82-84.
368. J.S. Schouten, F. DeJager y J.A. Greefkes, "Delta modulation, a New modulation system for telecommunication", *Phillips Technical Review,* vol. 13, pp. 237-245.
369. C.E. Shannon, "A mathematical theory of communication", *Bell System Tech. J.,* vol. 27, pp. 379-423, 623-656.
370. C.E. Shannon, "Communication theory of secrecy systems", *Bell System Tech. J.,* vol. 28, pp. 656-715.
371. C.E. Shannon, "Communication in the presence of noise", *Proceedings of the IRE,* vol. 37, pp. 10-21.
372. M.K. Simon y D. Divsalar, "On the irnplementation and performance of single and double differential detection schemes", *IEEE Trans. on Communications,* vol. 40, pp. 278-291.
373. B. Sklar, "A primer on turbo code concepts", *IEEE Communications Magazine,* vol. 35, pp. 94-102.
374. B. Sklar, "A structural overview of digital communications-A tutorial review", Parte 1, *IEEE Communications Magazine,* vol. 21, no. 5, pp. 4-17; Parte II, vol. 21, no. 7, pp. 6-21.
375. D. Slepian, "On bandwidth", *Proceedings of the IEEE,* vol. 64, pp. 292-300.
376. B. Smith, "Instantaneous companding of quantized signals", *Bell System Tech. J.,* vol. 36, pp. 653-709.
377. E.S. Sousa y S. Pasupathy, "Pulse shape design for teletext data transmission", *IEEE Trans. on Communications,* vol. COM-31, pp. 871-878.
378. S. Stein, "Unified analysis of certain coherent and noncoherent binary communication systems", *IEEE Transactions on Information Theory,* vol. IT-10, pp. 43-51.
379. C.E. Sundberg, "Continuous phase modulation", *IEEE Communications Magazine,* vol. 24, no. 4, pp. 25-38.
380. M. Tomlinson, "New automatic equaliser employing modulo arithmetic", *Electronics Letters,* vol. 7, pp. 138-139.
381. D.W. Tufts, "Nyquist's problem-The joint optimization of transmitter and receiver in pulse amplitude modulation", *Proceedings of the IEEE,* vol. 53, pp. 248-259.

382. G.L. Turin, "An introduction to matched filters", *IRE Transactions on Information Theory*, vol. IT-6, pp. 311-329.
383. G.L. Turin, "An introduction to digital matched filters", *Proceedings of the IEEE*, vol. 64, pp. 1092-1112.
384. G. Ungerboeck, "Channel coding with mutinivel/phase", *IEEE Transactions on Information Theory*, IT-28, pp. 55-67.
385. G. Ungerboeck, "Trellis-coded modulation with redundant signal sets", Partes 1 y 2, *IEEE Communications Magazine*, vol. 25, no. 2, pp. 5-21.
386. M.C. Valenti, "An introduction to turbo codes", EE Department, Virginia Polytechnic Institute & State University, Blacksburg, Virginia, sin publicar.
387. B. van der Pol, "The fundamental principles of frequency modulation", *Journal of* IEE (Londres), vol. 93, parte III, pp. 253-258.
388. J.H. Van Vleck y D. Middleton, "A theoretical comparison of visual, aural, and meter reception of pulsed signals in the presence of noise", *Journal of Applied Physics*, vol. 17, pp. 940-971.
389. A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Trans. on Information Theory*, vol. IT-13, pp. 260-269.
390. A.J. Viterbi, "Spread-spectrum communications-Myths y realities", *IEEE Communications Magazine*, vol. 17, no. 3, pp. 11-18.
391. A.J. Viterbi, "When not to spread spectrum-A sequel", *IEEE Communications Magazine*, vol. 23, no. 4, pp.12-17.
392. A.J. Viterbi, "Wireless digital communication: A view based on three lessons learned", *IEEE Communications Magazine*, vol. 29, no. 9, pp. 33-36.
393. G.K. Wallace, "The JPEG still picture compression standard", *Communications of the* ACM, vol. 34, pp. 31-44.
394. D.K. Weaver, Jr., "A third method of generation and detection of single-sideband signals", *Proceedings of the IRE*, vol. 44, pp. 1703-1705.
395. J. Weiss y D. Schremp, "Putting data on a diet", *IEEE Spectrum*, vol. 30, pp. 36-39.
396. T.A. Welch, "A technique for high performance data compression", *Computer*, vol. 17, no. 6, pp. 8-19.
397. L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space-part I: 180", *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, pp. 659-671.
398. L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space-part II: nonlinear codes", *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, pp. 672-686.
399. L.-F. Wei, "Trellis-coded modulation with multidimensional constellations", *IEEE Transactions on Information Theory*, vol. IT-33, pp. 483-501.
400. S.B. Weinstein, "Echo cancellation in the telephone network", *IEEE Communications Magazine*, vol. 15, no. 1, pp. 8-15.
401. S.B. Weinstein y P.M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform" *IEEE Transactions on Communications*, vol. COM-19, pp. 628-634.
402. J.J. Werner, "Tutorial on carrierless AM/PM-Part I: Fundamentals and digital CAP transmitter", *AT&T Bell Laboratories Report*, Minneapolis.
403. J.J. Werner, "Tutorial on carrierless AM/PM-Part II: Performance of bandwidth-efficient line codes", *AT&T Bell Laboratories Report*, Middletown.
404. B. Widrow y M.E. Hoff, Jr., "Adaptive switching circuits", *WESCON Convention Record*, Pt. 4, pp. 96-104.
405. J.H. Winters, "Smart antennas for wireless systems", *IEEE Personal Communications*, pp. 23-27.
406. J.H. Winters, "Adaptive antennas for wireless communications", Conferencia Internacional sobre Comunicaciones '99, Notas tutoriales TUS, Vancouver.
407. A.D. Wyner, "Fundamental limits in information theory", *Proceedings of the IEEE*, vol. 69, pp. 239-251.

408. J.L. Yen, "On the non-uniform sampling of bandwidth-limited signals", *IRE Transactions on Circuit Theory*, vol. CT-3, pp. 251-257.
409. O.C. Yue, R. Luganani y S.O. Rice, "Series approximations for the amplitude distribution and density of shot processes", *IEEE Transactions on Communications*, vol. COM-26, pp. 45-54.
410. N. Zervos e I. Kalet, "Optimized decision feedback equalization versus optimized orthogonal frequency division multiplexing for high-speed data transmission over the local cable network", Conferencia Internacional sobre Comunicaciones '89, pp. 35.2.1-35.2.6.
411. J. Ziv y A. Lempel, "A universal algorithm for sequential data compression", *IEEE Transactions on Information* Theory, vol. IT-23, pp. 337-343.
412. J. Ziv y A. Lempel, "Compression of individual sequences via variable-rate coding", *IEEE Transactions on Information Theory*, vol. IT-24, pp. 530-536.
413. W.Y. Zou y Y. Wu, "COFDM: An overview", *IEEE Transactions on Broadcasting*, vol. 41, pp. 1-5.

Notas

1. Las siguientes abreviaturas se utilizan en algunos de los artículos de revistas:

ACM: Association for Computing Machinery
AIEE: American Institute of Electrical Engineers
IEEE: Institute of Electrical and Electronics Engineers
IEE: Institution of Electrical Engineers (Londres)
IRE: Institute of Radio Engineers
SIAM: Society for Industrial and Applied Mathematics

ÍNDICE

A

acceso múltiple
  contra multiplexado, 513
  tipos básicos de, 513–514
acceso múltiple por división de código (CDMA)
  códigos para el, 505
  sistemas, 584
  ventajas del, 514
acceso múltiple por división de frecuencia (FDMA), 513, 516
acceso múltiple por división del espacio (SDMA), 514
acceso múltiple por división del tiempo (TDMA), 211–212, 513
  como sistemas de comunicaciones inalámbricos, 547–550
  sistema eficiente, 516
adaptable, 230
ADM. *Vea* modulación delta adaptable
adquisición, 493
ADSL. *Vea* líneas digitales asimétricas de suscriptores
Advanced Research Project Agency Network (ARPANET) e impacto en las comunicaciones por computadora, 28
"ajustadores de fase y de ganancia", 550
alfabeto de entrada, 582
alfabeto de salida, 582
algoritmo BCJR
  en función del algoritmo de Viterbi, 678
  exposición matemática de, 680
  formulación de, 678
  propósito de, 678
algoritmo de la modulación delta adaptable, 232
algoritmo de la transformada rápida de Fourier (FFT), 443–444
algoritmo de Lempel-Ziv, 8, 616
  comparado con la codificación de Huffman, 581
  definición de, 580
  estándar para la compresión de archivos, 581
  proceso de codificación efectuado mediante, 580
algoritmo de mínimos cuadrados, 288–290, 557
  ecuaciones para, 289
  para ecualización adaptable, 288–289
  para predicción adaptable lineal, 226
  popularidad, 226, 227
  resumen de, 289
  similitudes, 290
  simplificación, 289
  usos de, 292
  utilizando notación de matrices, 289
  y uso combinado, 297
algoritmo de mínimos cuadrados complejo
  limitaciones de, 558
  ventajas de, 558
  Vea también algoritmo de mínimos cuadrados
algoritmo de suma-producto, 691
algoritmo de Viterbi, 661–663, 670
  como un decodificador de probabilidad máxima, 662
  como un estimador de secuencia de probabilidad máxima, 662
  dificultad en la aplicación de, 662
algoritmo del estándar de encriptado de datos (DES), 754
algoritmo recursivo
  para recuperación de fase, 454
  para recuperación de la temporización, 457–458
algoritmo recursivo de decisión dirigida, 450
algoritmo recursivo sin asistencia de datos, 455
algoritmo Rivest-Shamir-Adleman (RSA), 759
algoritmo RSA, 757–758
algoritmos de compresión estándar, 8
algoritmos de decodificación, 678
alisamiento, 223
AM. *Vea* modulación de amplitud
ambiente de las telecomunicaciones, 3–7
ambiente de multitrayectoria "dinámico", 532–533
ambiente DSL, 447
American Standard Code for Information Interchange (ASCII), 6, 762
amplificación, 128
amplificador de tubo de onda viajera, 516
amplitud del pulso transmitido, 253
amplitud/fase sin portadora (CAP), 369, 380
análisis de enlace por radio, 517–523
análisis de Fourier, 715–720
análisis de modulación por tonos, 119
análisis de ruido
  comparación de, 163–164
  en el receptor, 523
análisis del espacio de la señal, 337
análisis del presupuesto del enlace, 517
análisis espectral, 110
análisis espectral de dos etapas, 110
analizador de ruido de banda angosta, 64–66
ancho de banda, 720–723
  definiciones de, 720–721
  eficiencia de, 347, 348
ancho de banda de 3 dB, 721
ancho de banda de coherencia, 540
ancho de banda de la señal, 3
ancho de banda de nulo a nulo, 368, 721
ancho de banda de Nyquist, 262
ancho de banda de transmisión
  definición, 118
  dispersión instantánea, 499

ancho de banda de transmisión
normalizada, 164
ancho de banda de video, 6
ancho de banda del canal
definición de, 3
fuente de comunicación principal, 92
ocupación de, 347
utilizado en Estados Unidos, 102
ancho de banda del mensaje, 90
ancho de banda del valor cuadrático medio (rms), 721
ancho de banda equivalente del ruido 722–723
definición, 723
para filtros pasobanda, 723
ancho de banda infinito, 602
ángulo, 107, 312
antena
ancho del haz, 521
diseño de, 17–18
extremo de recepción, 18
uso de haz múltiple, 514
antena de haz múltiple, 514
antena de referencia
como una fuente isotrópica, 519
definición de, 519
antena de transmisión
como fuente puntual, 519
función de, 518
montaje de, 17–18
y densidad de potencia, 521
antena receptora, 518
antena reflectora, 522
antenas de haces múltiples, 514
anulaciones espectrales, 368
apertura efectiva, 521
apertura ocular, 293
apiñamiento de bits, 215
aproximación lineal por secciones, 203
apuntador, 581
árbol de código, 657, 657–658
árbol de decisión, 575–576
árbol de fase, 388
Armstrong, Edwin H., 27
ARQ. *Vea* solicitud de repetición automática
arquitectura de Internet, 13–14
arquitectura en capas, 11
arreglo de antena adaptable, 557
como técnica especializada, 559
estructura de, 557
optimización del desempeño de, 558
uso de, 559
arreglos de antenas, 553
artículo memorable de Shannon, 567
ASCII. *Vea* American Standard Code for Information Interchange línea digital
asimétrica de suscriptor (ADSL)
motivación para, 282
servicios soportados por, 281
uso del, 446
ventajas de, 446
ATM. *Vea* modo de transferencia asíncrona
AWGN. *Vea* ruido blanco gaussiano aditivo

B

banda base, 88
banda crítica, 234
banda de frecuencia asignada, 154–155
banda de frecuencia intermedia (FI), 18
banda lateral, superior e inferior, 91
bandas de seguridad, 513
Bardeen, John, 28
base, 570, 576–577
Baudot, Emile, 26
bauds, 276
Bayes, regla de, 585–587, 707
Bell, Alexander Graham, 27
Berners-Lee, Tim, 28–29
B-ISDN. *Vea* red digital de servicios integrados de banda ancha
bit, 204, 569
bit de detección de errores, 7
bit de paridad, 7, 632
bits de datos
para detección de errores, 7
patrón binario de, 6
bloques, 572
bloques funcionales de, 13–14
Brattain, Walter H., 28
brecha, 432
brecha de energía, 99
brecha de la relación señal a ruido, 432
brinco lento de frecuencia, 500–502
British Broadcasting Corporation (BBC), 27
byte, 7

C

cable coaxial
aplicación de, 17
componentes de, 17
contra pares trenzados, 17
cable de par trenzado
componentes de, 16
contra cables coaxiales, 17
susceptible a, 16–17
usos de, 277
calculadora del síndrome, 647
cálculos de ruido, 61
calidad de reproducción, 5–6
calidad de servicio, 14
cama de un tiempo, 747
campo entrelazado, 5
canal, 2
características del, 309–310
respuesta en frecuencia de, 261
y ruido aleatorio, 259
canal AWGN continuo, 318–319
canal con ruido blanco gaussiano aditivo (AWGN) de entrada binaria, 667, 668
canal de ancho de banda limitado, 16
canal de banda base
requerimientos del canal, 247
y transmisión de datos digitales, 247
canal de comunicación, 7, 15–19, 277
clasificación de, 3, 15, 19
descripción de, 15–19
uso de, 88
uso simultáneo de, 512
canal de comunicación pasobanda, 348–349
canal de desvanecimiento
características de, 541
efectos de, 545
canal de desvanecimiento de Rayleigh, 536, 541
desempeño de, 545
sobre señalización binaria, 542–547
canal de desvanecimiento de Rician, 536
canal de desvanecimiento de trayectoria múltiple, 536–539
canal de desvanecimiento plano de Rayleigh, 554
canal de desvanecimiento plano, 71–72
canal de Nyquist ideal, 262–264, 265–266
dificultades de, 263–264
uso de, 263–264
canal de potencia limitada, 3
canal de radio móvil
capacidad de, 18

como un canal lineal variante con el tiempo, 18
efectos de propagación de, 18
canal de ruido blanco gaussiano aditivo (AWGN), 559
capacidad, 431–432
caracterización de, 322
señal recibida para, 403
técnicas en, 464
y detección de señal, 329
y probabilidad de error, 332
y transmisión de señal, 309
canal de ruido coloreado, 607–611
canal de satélite
accesos a áreas remotas, 18
capacidades de, 516
cobertura de, 18
canal de trayectoria múltiple
caracterización estadística de, 535–542, 554
clasificación de, 541–542
como un deterioro del canal, 553
modelo de, 71–72
selección de frecuencia, 541–542
tipo de desvanecimiento exhibido, 72
canal discreto sin memoria, 581–584, 629–631
capacidad de canal de, 588
definición, 581–582
canal doblemente disperso, 542
canal en cuadratura, 408–409
canal en fase, 408
canal en tiempo discreto, 291
canal gaussiano, 597
canal gaussiano sin memoria, tiempo discreto, 597
canal *I*, 97
canal limitado en banda, 3
canal no plano, 542
canal plano de frecuencia fija, 542
canal plano en el tiempo, 542
canal plano-plano, 542
canal Q, 97
canal seguro, 743
canal simétrico binario (BSC), 258, 629, 667
canal sin memoria, 321
canal telefónico, 15–17, 287
canal transpondedor, 515
canales de comunicación de banda ancha, 218
canales de transmisión inalámbrica, 17–18
cancelación del eco, 277–278
comparación de esquemas, 278–279
modo de operación para, 277–278
cancelador del eco
en tranceptores, 278–279
uso de, 516
CAP. *Vea* amplitud/fase sin portadora
capa, 11
capa física, 13
capacidad, 568
capacidad de información, 598, 616
de un canal, 597–598
definición, 23, 598
evaluación de, 598
incremento de, 599
capacidad de información óptima, 610
capacidad del canal, 587–589
concepto de, 630–631
de canal discreto sin memoria, 588
capacidad del sistema
definición de, 549
descripción de, 718
para maximizar, 549
característica de caída del coseno completo, 266
característica gaussiana, 75
características de antiperturbación, 498–499
CDMA. *Vea* acceso múltiple por división de código
celdas, 14–15
celdas de cocanal
búsqueda de, 531
como interferencia, 553
determinación de, 531
CELP
característica distintiva de, 552
decodificador para, 552
implementación de codec, 553
modelado de, 616
Vea también LPC de código excitado
centro de conmutación móvil, 530
cero de señal digital (DS0), 214
cifra de ruido, 523–526
cifrador, 742
cifrador de bloque, 744, 745, 746
cifradores de frecuencia aditivos binarios, 744
circuitería de compresión y expansión, 203
circuito, 10–11
circuito de pendiente, 121–122
circuito de pendiente ideal
caracterizado por, 121–122
respuesta en frecuencia de, 122
circuito de realimentación de multilazo, 480
circuito de recuperación de fase, 345
circuito equivalente de Thévenin, 60
circuito equivalente Norton, 60
circuito fotodetector, 58–59
circuitos integrados a muy grande escala de integración (VLSI)
desarrollo de, 28
Clark, Arthur C., 29
clave de producto, 749
clave de sustitución
descripción de, 749–750
uso de, 749–750
clave de transposición
descripción de, 750
uso de, 750
clave de Vernam, 747
claves en flujo, 744–746
operación de, 744
utilizadas en, 745
clic que se vuelve negativo, 150
clic que se vuelve positivo, 150
cobertura mundial, 512
codec, 552
codec de análisis por síntesis, 552
codificación convolucional, 669
codificación de control de errores
clases de, 626
para comunicación confiable, 567
técnicas, 626
técnicas para, 683, 693
uso de, 626, 627
codificación de decisión dura, 630
codificación de enramado
como un esquema de corrección de errores directos, 424
como una técnica de codificación de control de errores, 430
codificación de fuente, 574
con un criterio de fidelidad, 611–612
disección de, 616–617
para comunicación eficiente, 567
codificación de Huffman, 578–580, 616
como codificación entrópica, 8
comparada con el algoritmo de Lempel-Ziv, 581
idea básica de, 578
y compresión de datos, 8
codificación de Lempel-Ziv, 580–581
codificación de nivel correlativo 266–271

base de, 268
forma generalizada de, 274–275
idea ilustrada 267
premisa de la, 267
codificación de percepción, 9
codificación de pulsos
en modulación analógica, 217
en modulación digital por pulsos, 183
uso de, 184
codificación del canal
meta de diseño de, 589
técnicas para, 628
y mapeo, 589
codificación del habla
a bajas velocidades, 229–230
a bajas velocidades de bit, 229–230
aplicaciones de, 229–230
filosofía del diseño de, 230
técnicas para, 551
codificación del prefijo, 575
codificación directa de control de error, 628, 668
codificación duobinaria, 270
codificación multinivel, 348, 378
codificación no redundante, 421
codificación predictiva lineal (LPC), 551
codificación turbo, 674
codificador
condición para la optimización de, 199–200
estados de, 659, 667
operación de, 646
partes principales, 551
unidades funcionales de, 235
codificador binario doblemente modificado
característica útil de, 272–273
respuestas de, 272, 273
codificador convolucional
árbol de código para, 657–658
enramado para, 658–659
estado de, 657–658
relación de entrada-salida de, 660
uso de, 425
codificador de constelación, 444
codificador de enramado, 425
codificador de fase diferencial, 415
codificador de fuente, 21, 574
propósito de, 21
requerimientos funcionales de, 574
codificador de nivel activado-desactivado, 384
codificador de nivel polar sin retorno a cero (NRZ), 352
codificador de transmisión de la señal, 348, 352
codificador diferencial
componentes de, 421–422
de la onda binaria, 414
método utilizado, 207
requerimiento de, 207
codificador duobinario, 267–268
codificador Gray, 276
codificador sin regreso al nivel cero, 359
codificador turbo, 683
código 2B1Q, 281
como el estándar estadounidense, 281
comparado con otros códigos de línea, 281
niveles de amplitud de, 317
propiedades deseables de, 281
código algebraico
propiedades de, 641–642
tipos de, 693
código BCH binario
contra no binario, 654
tipos comunes de, 653
código binario
eficiencia de, 197
símbolos de, 204
código binario doblemente modificado, 281
código bipolar, 281
código catastrófico, 667
código cíclico, 641–643
características de, 652–654
clases de, 652–654
decodificador para, 645–646
en forma sistemática, 645, 646
generación de, 643
propiedades de, 642
ventaja de, 641–642
código cíclico de verificación de redundancia (CRC)
generador de polinomios de, 653
para detección de errores, 652
código convolucional, 654–656
característica distintiva de, 627
decodificación de máxima verosimilitud del, 660–663
desempeño de, 663
longitud de restricción de, 655
propiedades de distancia de, 663
uso de, 654–656
código convolucional simétrico recursivo (RSC), 675
código CRC. *Vea* código cíclico de verificación de redundancia (CRC)
código de bloque, 590, 632
característica distintiva de, 627
velocidad de, 685
código de bloque sistemático, 632
código de control de errores
teoría de, 485
tipos de, 627
código de corrección de error simple de Hamming, 653
código de Huffman
algoritmo utilizado para sintetizar, 578
como una clase de códigos de prefijos, 578
desventaja de, 580
no unicidad de, 579
código de línea pasobanda, 344
código de longitud variable, 574
código de Manchester, 207, 281
código de perforación, 676
código de prefijo
distinción mediante, 575–576
distinguido por, 577
propiedad de, 576
código de prefijo extendido, 578
código de verificación de paridad de baja densidad (LDPC), 683–686
algoritmo de decodificación, 690
asignación de valores iniciales de, 690
construcción de, 684–685
decodificación de, 689–690
distancia mínima de, 689
longitud de bloque de, 689
ventajas de, 684
código disperso
con propiedades pseudoaleatorias, 490
uso de, 490
código doble, 641
código duobinario modificado, 281
código extendido
longitud promedio de palabra de código, 578
uso de, 577
código fuente, 574
tipos de, 575
variabilidad en longitudes de, 579
código irregular turbo, 691, 692
código LDPC irregular, 692

código Morse, 26, 574
código no sistemático, 656
código probabilístico, 693
código RS de corrección del error de *t*, 654
código telegráfico de Baudot, 26
código ternario, 204
código turbo ordinario, 692
códigos adecuados, 631
códigos BCH primitivos, 653
códigos BCH, 653–654
códigos Bose-Chaudhuri-Hocquenghem (BCH), 653–654
códigos compuestos, 683–684
códigos CRC binarios
  capacidades de, 652
  y detección de errores, 652–653
códigos de bloque lineal
  arreglo estándar de, 638
  clases de, 653–654
  definición de, 632
  distancia mínima de, 637
  estructura matemática de, 632–633
  procedimiento de decodificación para, 639
  propiedad básica de, 634
códigos de enramado
  diseño de, 669
  para canales de banda limitada, 668–669
códigos de línea
  candidatos para, 281
  comparación de, 281
  espectros de potencia de, 206
  para representación eléctrica, 204–207
  selección de, 280–281
  tipos de, 205–207
  uso de, 204–207
Códigos de Ungerboeck, 670
  ganancia de codificación asintótica de, 673
  para PSK 8, 670–672
códigos instantáneos, 577
códigos irregulares, 691
códigos Reed-Solomon, 654, 693
códigos turbo
  componentes de, 674
  desarrollo de, 674
  desempeño de, 676–677
  procedimiento de terminación de, 676
  propiedades de, 682, 693
coeficiente de correlación, 342, 471
coeficiente de Fourier complejo, 717
coeficiente de la transformada coseno discreta, 8
colores primarios, 6
  representados mediante señales de video, 6
  transmisión de, 6
combinador de relación máxima, 547
combinador lineal, 547
como una disciplina matemática, 567
compactación de datos, 8, 575
  logro de, 575
  esquemas para, 575
  evaluación, 616
compander, 203
complejidad de la decodificación, 684
componente de la portadora, 135
componente de ruido en fase, 131
componente de trayectoria múltiple, 549
componente en cuadratura, 93
  densidad espectral de potencia de, 386
  papel de, 93, 101
  propiedades de, 65–66
componente en fase, 93
  densidad espectral de potencia de, 395–396
  propiedades de, 65–66
  representación de, 67–69
componentes de ruido en cuadratura, 131
compresión con pérdidas
  definición de, 8
  el método preferido, 8
compresión de datos, 7, 614–616
  como una operación con pérdidas, 614–615
  componentes del sistema, 749
  en criptografía, 749
  formas de, 7–8
  idea de, 614
  razón para su uso, 615
  técnicas para, 218
compresión de datos con menores pérdidas, 575
compresión sin pérdidas
  contra compresión sin pérdidas, 8
  definición de, 7–8
  para texto digital, 8
compresor, 203
compresor de datos, 614
compromiso ancho de banda-ruido, 193
compromisos, 602
computadora de arco tangente, 364
computadora de diferencia de fase, 364
computadoras personales (PC), 6, 6–7
comunicación, 1
  aplicaciones, 1
  frecuencias para, 4
  fundamento de, 2–3
  tipos de, 21–23
comunicación de datos, 7
comunicación de voz, 22
comunicación digital
  confiabilidad de, 23, 24–26
  elementos de la, 24–25
  forma básica de, 309–310
  receptor, 337
  requerimientos de, 23
  tarea del diseñador, 626
  uso de, 21
  y diseño de sistemas, 22
  y objetivos de diseño, 354
  y tasa de error de bits, 24
comunicación digital por satélite, 419
comunicación humana, 4
comunicación óptica, 29
comunicación punto a punto, 2–3
comunicación sin errores, 568
comunicación virtual, 13
comunicaciones alámbricas, 559–560
comunicaciones inalámbricas, 529–535
  arreglos de antenas adaptables para, 553
  características de, 512
  codificación de fuente para, 550–553
  como un tipo de comunicación multiusuario, 512
  como un tipo de sistema de comunicación de radio multiusuario, 529–530
  contra comunicaciones alámbricas, 559–560
  objetivo de, 553
  principales deterioros de canales de, 553
  y movilidad, 512
  y OFDM, 448
comunicaciones multiusuarios, 512
  ambiente, 396
  tipos de 559
comunicaciones por computadora, 2
comunicaciones por satélite, 514–517
  banda de frecuencia más popular para, 515
  banda de frecuencia para, 19

cobertura global, 512
como un tipo de comunicaciones multiusuario, 512
comunicaciones seguras
en un ambiente hostil, 479
necesidades de, 742
concepto celular, 530
conceptos probabilísticos, 703–707
conceptos teóricos de la información, 572
condición del prefijo, 575
condición del vecino más cercano, 200
condiciones de Dirichlet, 715
conductor exterior, 17
conductor interior, 17
conexión de larga distancia, 16
conexión en cascada
de redes de dos puertos, 524–525
figura de ruido de, 526
configuración del módem, 421
configuraciones simétricas del módem, 421–425
confusión, 749
conmutación a bordo, 516
conmutación de circuito, 10–11
conmutación de paquetes
principio de, 11
red, 11
conmutador, 211
conmutador de Strowger, 27
conservación de tiempo, 211–212
constelación circular, 368
constelación de las señales, 322–323, 337
como simetría circular, 335
construida a partir de símbolos PCM unidimensionales, 429
definición de la distancia mínima de, 335
constelaciones al cuadrado, 369–370
constelaciones cruzadas, 369–370, 372
constelaciones de modulación de amplitud en cuadratura (QAM), 369–370
contenido espectral, 492, 493
continuidad de fase, 388
control de errores, 626
control de potencia
en sistemas CDMA, 549
uso de, 549
conversión ascendente, 104–105
conversión de AM a PM, 127–128
conversión descendente, 105
convertidor analógico-digital (A/D), 201, 379–380, 445
convertidor ascendente, 448
convertidor ascendente de frecuencia, 105
convertidor descendente, 448
convertidor descendente de frecuencia, 105, 516
convertidor digital/analógico (DAC), 445
convertidor paralelo/serie, 445
convertidor serie a paralelo, 445
convolución en tiempo discreto, 627
corrección de errores directa (FEC), 626, 628, 629
corrección de fase, 365
correlator, 24–25
entradas de, 549–550
salidas de, 319–322
corriente de salida, 4–5
corrimiento de amplitud por llaveo (ASK)
corrimiento de cuadrifase por llaveo (QPSK), 354, 354–361
caracterización de, 354–355
diagrama del espacio de la señal, 354–355
modo de operación, 425
motivación para utilizar, 508
probabilidad de error de, 356–358
corrimiento de fase por llaveo (PSK), 24
esquema de señalización, 344–345
esquemas de corrimiento de fase por llaveo (PSK), 418
modulación de fase, 345
representada por medio de, 464
y sistemas coherentes, 490
corrimiento de fase por llaveo (PSK) binaria coherente
características, 350
tasa de error de bits, 352, 417
corrimiento de fase por llaveo (PSK) binaria, 349–353
como operación lineal, 492
modulador, 490
probabilidad de error de, 350–352
señales, 353
transmisor, 352
uso de, 550
y modulación de espectro disperso, 550
corrimiento de fase por llaveo (PSK) coherente, 349
corrimiento de fase por llaveo diferencial (DPSK) binaria, 415
corrimiento de fase por llaveo diferencial (DPSK), 407, 414–417
generación y detección de, 415
receptor, 416
tasa de error de bits de, 417
transmisor, 415–416
corrimiento de fase por llaveo M-aria de secuencia directa (DS/MPSK), 508
corrimiento de fase relativa, 532
corrimiento de frecuencia por llaveo (FSK)
esquema de señalización básica, 344–345
representada por, 464
y diseño de módems, 421
y modulación de frecuencia, 345
corrimiento de frecuencia por llaveo (FSK) binaria
probabilidad de error de, 382–384
tasa de error de bits para, 384
corrimiento de frecuencia por llaveo (FSK) binaria coherente, 380
corrimiento de frecuencia por llaveo (FSK) binaria no coherente, 407, 413–414
consideraciones de, 413
expresiones para, 417
tasa de error de bits para, 414
corrimiento de frecuencia por llaveo en fase continua (CPFSK), 381, 387
corrimiento de frecuencia por llaveo M-aria (MFSK), 398–400
componentes de, 401
para sistemas de salto de frecuencia, 500
propiedad de, 400
corrimiento de frecuencia por llaveo M-ario de salto de frecuencia (FH/MFSK), 508
corrimiento de octafase por llaveo, 365–366
corrimiento Doppler, 535
corrimiento Doppler medio, 541
corrimiento en cuadrifase por llaveo (QPSK) coherente, 354
corrimiento mínimo por llaveo (MSK), 387
con una forma de FSK binaria, 361
diagrama del espacio de la señal de, 389–392

corrimiento mínimo por llaveo de filtrado gaussiano (GMSK)
característica indeseable de, 398
como un tipo especial de modulación de frecuencia binaria, 398
cota de Cramér–Rao
definición, 462
modificación de, 462
cota de unión, 337
como una cota superior útil, 332–333
ilustración de, 333
simplificación de, 335
uso de, 401
cota fundamental de Shannon para la seguridad perfecta, 747
cota superior, 401
cotas
para predicción, 332
uso de, 332
CPFSK. *Vea* corrimiento de frecuencia por llaveo en fase continua
crecimiento de, 29
criptoalgoritmo de llave secreta, 751
cripto-algoritmo RSA, 758
criptoanálisis, 742
definición de, 746
descripción de, 742
y usuario autorizado, 742–743
criptoanalista enemigo
formas de ataque de, 745
intromisión de, 743, 744
criptografía, 617, 742
clasificaciones de, 759
compresión de datos en, 749
importancia de, 759
suposición fundamental, 745
y el problema de autentificación, 743
y el problema de privacía, 743
criptografía de doble llave, 742
criptografía de llave pública, 742, 755–757
criptografía de llave secreta, 742–743
criptografía de una sola llave, 742
criptograma, 742–743
criptología, 742
criterio de Nyquist, 262
criterio de toma de decisiones, 323
criterio del error cuadrático medio
para el diseño de receptor, 283
usos de, 288
cuantización adaptable, 230
con estimación directa, 230
con estimación inversa, 230
problemas de, 230
cuantización
aplicación, 202–203
función de, 196
propósito de, 8
tipos de, 194–195, 220
uso de, 195
y codificación, 9
cuantización de amplitud, 194
cuantización escalar
forma de, 194
uso de, 194
cuantizador
características de, 194
clases de, 615
como un compresor de señal, 615
componentes de, 199
tipos de, 194
cuantizador adaptable, 230
cuantizador de Lloyd-Max, 198, 200–201
cuantizador de vectores,
contra cuantizador escalar, 615
proceso de codificación en, 615
relación señal a ruido de cuantización para, 615
ventajas de, 615
cuantizador escalar
como un compresor de señal simple, 615
condiciones para la optimización de, 198–201
diseño de, 198
cuantizador no uniforme, 202–203
cuantizador sin memoria, 194
cuantizador uniforme, 196
cuantizador, no uniforme, 202–203
curva universal, 118–119

D

datos binarios NRZ, 398
datos generados por computadora, 7
decisiones de decodificación, 663
decisiones de fase, 394
decisiones duras, 630
decisiones flexibles, 630
decodificable singularmente, 574
decodificación de decisión flexible, 669
decodificación de Huffman, 8–9
decodificación del síndrome, 635–636, 638, 639
decodificado turbo, 677–680, 682–683
decodificador, 446
componentes de, 552
condición para optimización, 200–201
función de, 552
unidades de, 235
decodificador algebraico, 629–630
decodificador de canal
bajo el control del diseñador, 590
objetivo de, 627
operación de correspondencia inversa, 590
decodificador de distancia mínima, 661
decodificador de fuente, 575
decodificador de máxima verosimilitud, 322–326
como un dispositivo de implementación, 324
definición, 661
para cálculo, 324
teoría de, 660–661
decodificador de transmisión de la señal, 326, 349
decodificador del canal, 626
bajo el control del diseñador, 590
introducción de redundancia, 590
objetivo del, 627
operación de correspondencia, 590
suposición markoviana para, 681
decodificador MAP, 678
decodificador probabilístico, 630
decodificador turbo
complejidad de, 684
detalles de, 683
estructura básica de, 677–678
decodificadores de decisión dura, 629–630
de Forest, Lee, 27
demodulación, 20, 88
etapas de, 491
método de, 99
demodulación de decisión dura, 669
demodulación de fase, 492
demodulación de frecuencia
definición, 121
métodos de, 121
demodulador de FM
con realimentación negativa, 153
y tipo de oscilador, 152
demodulador de malla de fase encadenada, 152
como un filtro de rastreo, 154
y capacidad de extensión del umbral, 154
demodulador de pulsos, 211

demodulador FMFB, 152, 153, 154
demoduladores de umbral extendido, 152, 153
demoduladores individuales, 106
demultiplexor
  en receptor, 359, 445–446
  en transmisor, 444
densidad espectral de energía, 48, 353
densidad espectral de potencia de banda base de una señal de PSK binaria, 353
densidad espectral de potencia, 44–46, 347
  de proceso aleatorio, 50, 52
  importancia de, 45
  porciones de frecuencia de, 155
  propiedades de, 46–47
  resumen gráfico de, 75–76
  y espectro de amplitud, 50–52
densidades espectrales cruzadas, 52
densidades marginales, 710
DES. *Vea* detección del estándar de encriptado de datos
  de una señal de pulso, 248
  y corrección de errores, 628
desatendido, 530
descomposición espectral, 443
desconmutador, 211
descripción en el dominio de la frecuencia, 444, 720
descripción en el dominio del tiempo, 720
desencriptado, 742
desigualdad de Chebyshev, 713
desigualdad de Kraft-McMillan, 576–577
desigualdad de Schwarz
  como un resultado matemático, 249–250
  demostración, 313, 314
desigualdad fundamental, 571
desvanecimiento de la tasa de bits, 218
desvanecimiento de señal, 532–533
desvanecimiento del umbral, 152
desvanecimiento del umbral de FM, 152–154
desviación de frecuencia, 110, 152
detección coherente, 98, 131, 133
  efecto de, 97–98
  uso de, 132
  y demodulación, 95
detección de envolventes, 102–103, 131
detección de máxima verosimilitud, 330, 337
detección de probabilidad *a posteriori* (MAP) máxima, 678
detección de señal de máxima verosimilitud, 346
detector, 326, 349
detector coherente de fase en cuadratura, 97
detector coherente en fase, 97
detector de envolvente, 123, 143
  comparación de señales, 141–142
  componentes de, 92
  desempeño de, 137
  encontrados en, 92
  necesidad de, 406
  pérdida del mensaje en, 138
detector de envolvente ideal, 135
detector diferencial
  componentes de, 364
  tipo tangente, 364–365
detector FSK M-ario no coherente, 500
detectores de máxima verosimilitud, 435
DFT. *Vea* transformada discreta de Fourier
diafonía, 21
  causa de, 279
  como deterioro, 279
  definición, 501
  tipos de, 279–280
diafonía del extremo cercano (NEXT), 279–280, 380
diafonía del extremo lejano (FEXT), 279–280
diagrama de estado, 657–660, 659
dibits, 276
dibits de codificación Gray, 363
diferencia de fase relativa, 414
diferenciador, 143
difracción, 17–18
difusión, 749
dígito binario, 204, 569
dimensionalidad del espacio de la señal, 312, 493
direccionalidad, 520
discriminador, 144
discriminador de fase, 97
discriminador de frecuencia ideal, 124
discriminador de frecuencia, 121, 124
  componentes de, 121–122
  entrada, 149
  requerimientos de, 99
diseñador de sistemas de comunicación, 23
dispersión del espectro
  como una operación lineal, 492
  y modulación de fase, 491
dispersión del retardo, 539
  como un deterioro del canal, 553
  definición, 540
  efecto de, 556
dispersión Doppler, 539–541
dispersores, 71–72
dispersores fijos, 536
dispositivo de decisión, 259
dispositivo de toma de decisiones
  diseño de, 277
  operación de, 25
distancia de Hamming, 637, 661, 666
distancia de unicidad, 748
distancia libre, 663
distancia mínima
  consideraciones de, 637–638
  definición de, 637
distorsión, 2
  aceptable, 614
  inevitable, 611–612
  métodos de desvanecimiento, 103
  producción de, 102–103
distorsión armónica, 112
distorsión de amplitud, 191
distorsión de amplitud, 191
distorsión de envolvente, 90–91
distorsión de fase
  presencia de, 99–100
  y el oído humano, 99–100
distorsión de la forma de onda, 102–103
distorsión de la media cuadrática, 199
distorsión de sobrecarga de pendiente, 220, 221
distorsión máxima, 288
distribución de Poisson, 59
distribución de probabilidad conjunta, 583
distribución de probabilidad marginal, 583
distribución de probabilidad, 583
distribución de Rayleigh, 68–69, 70, 74–75
distribución de Rician
  forma normalizada de, 71
  presentación gráfica de, 70–71
distribución del espectro
  como una operación lineal, 492
  en la demodulación, 491

distribución gaussiana, 54, 72–73
distribución gaussiana normalizada, 54–55, 56
diversidad de frecuencia, 544–545
diversidad de tiempo, 544–545
diversidad del espacio, 544–545
"divide y vencerás", 431
división, 488, 501
división de celda, 531
división del canal en tiempo continuo, 432–436
divisor, 282
DMT. *Vea* multitono discreto
DPSK binaria no coherente, 407
DPSK. *Vea* corrimiento de fase por llaveo diferencial
duplexado por división de frecuencia (FDD), 547
duración de bit, 253
duración de división, 494
duración rms, 722

E

ecuación de análisis, 442–443
ecuación de Bessel, 735
ecuación de Bessel modificada, 738
ecuación de Friis del espacio libre
  definición, 522
  usada para, 522
ecuación de síntesis, 442–443
ecuación del generador, 634
ecuación diferencial de Kummer, 740
ecuaciones de verificación de paridad, 634
ecuaciones de Wiener-Hopf, 224
ecualización adaptable
  algoritmo para, 287–288
  como método de selección, 379
ecualización con realimentación de decisiones, 291–293, 379, 430
ecualización lineal, 379, 556
ecualizador, 191
ecualizador adaptable
  capacidad de seguimiento de, 291
  método preferido de, 297
  modos de operación para el, 290
ecualizador con realimentación de decisiones (DFE)
  componentes de, 292
  propagación de errores en, 292–293
  sección de alimentación directa de, 292
  sección de realimentación de, 292
ecualizador de forzamiento a cero, 283
ecualizador espaciado fraccionalmente (FSE), 287
ecualizador óptimo, 288
ecualizador síncrono adaptable, 287
ecualizador transversal, 286
efecto de apertura, 191
efecto de captura, 148–149
efecto de umbral, 137–138
  AM y FM, 164
  clics escuchados en, 149–150
  definición de, 138, 149
  en un detector de envolvente, 138
efecto de umbral de FM, 149–152
efecto de voz del pato Donald, 99–100
efecto nulo de cuadratura, 97
efecto nulo de cuadratura de demodulación síncrona, 96
efecto ruido-silenciamiento, 147
efectos de la propagación, 532–535
eficiencia de la codificación, 574
eficiencia del ancho de banda, 347, 348
  definición, 347
  diagrama de, 601–602
  producto de, 348
EIRP. *Vea* potencia efectiva radiada referida a un almacenamiento elástico de fuente isotrópica, 215
elemento de retardo ideal, 268
elemento observable, 321
elementos de código, 203–204, 212
elementos de retardo unitario, 646
empaque de tramas, 9
empaquetamiento en esfera, 599–600
en el dominio de la frecuencia, 251
encriptado, 742
encriptado de datos, 617
energía por bit de la señal transmitida, 258
energía promedio mínima, 332
ENIAC, 28
enlace ascendente, 19, 514–515
enlace de comunicación
  análisis de, 517
  en conmutación de circuitos, 11
enlace de comunicación por radio, 522
enlace de modulación por codificación de pulsos (PCM), 210
enlace descendente, 19, 514–515
enlace directo, 547
enlace dúplex completo, 628
enlace dúplex medio, 628
enlace inverso
  contra enlace directo, 559
  subbandas, 547
enlaces fijos de punto a punto, 18–19
enmascaramiento auditivo, 9
enmascaramiento de ruido, 234
enramado, 657–660, 661–662
enramados de fase, 388–389
entrada de canal fija, 582
entrada del canal, 597, 610
entrada del cuantizador, 221
entrada flexible-salida flexible, 693
entrada flexible-salida rígida, 693
entropía absoluta, 594
entropía diferencial, 593–597
entropía
  condicional, 584
  definición de, 569
  fórmula para, 568
  propiedades de, 570–571
envolvente
  definición, 730
  tipos de, 730
  y componentes de fase, 67–69
envolvente compleja, 347, 727, 728, 729, 730
envolvente constante, 111
error
  minimización del, 551
  tipos posibles de, 254
  probabilidad de, 497–499
error acumulativo, 221–222
error cuadrático medio de exceso, 291
error cuadrático medio mínimo (MMSE), 558
  criterio para, 558
  ecualizador, 285
  receptor, 286–287
error cuadrático medio, 285
  como la función de costo, 558
  definición de, 284
error de decodificación, 660
error de fase
  definición, 158
  efecto de, 99
error de símbolo
  probabilidad condicional de, 333
  probabilidad promedio de, 276
error de temporización, 264
errores de decisión, 210
errores sucesivos, 232
escaneo, 4–5, 4–6
escaneo de trama de entrelazado, 5
escaneo de trama, 4–5

escaneo secuencial
de imágenes, 4
proceso de, 4–6
esferas de decodificación
número máximo de, 600
empaquetamiento de, 600
espacio de muestreo, 32, 704
espacio de observación, 324, 325–326
espacio euclidiano *N*-dimensional
ángulos en, 312
longitudes de vectores, 312
vectores en, 311, 312
espacio vectorial, 554–555
espectro, 3, 4, 715
espectro de fase, 715
espectro de magnitud, 715
espectro de potencia, 4, 45
espectro de potencia de retardo, 539
espectro de traslape, 187
espectro de voz, 3
espectro del coseno elevado
parte caída de, 264–265
parte plana de, 264–265
espectro del mensaje
para frecuencias negativas, 103
requerimiento de, 99
espectro disperso
atributo importante de, 488
comunicaciones, 508
noción de, 488–490
espectro disperso de salto de frecuencia, 499, 500
principios de, 480
sistemas de comunicación, 500
espectro disperso de secuencia directa
con BPSK coherente, 490–493
principios de, 480
sistemas, 498
espectro Doppler, 540–541
espectros de potencia, 347
esquema binario doblemente precodificado, 270–271
esquema de codificación en paralelo, 676
esquema de codificación Gray, 422–423
esquema de cuantización diferencial, 227, 228
esquema de demodulación, 132
esquema de filtrado, 100
esquema de modulación fijo, 627
esquema de modulación, conservación del ancho de banda, 354
esquema de señalización básico, 344–345
señales, 345
esquema de señalización de respuesta parcial, 267
características útiles de, 275
clases de, 275
logro de, 274–275
esquema de señalización duobinario, 267–271
respuesta en frecuencia de, 268
técnica, 271–272
esquema de señalización *M*-aria, 345–346
esquema QPSK desfasado p/4, 363
esquemas de corrimiento de frecuencia por llaveo (FSK), 418
esquemas de modulación de fase, 368
esquemas de modulación digital
comparación de, 417–420
empleando una sola portadora, 417–420
probabilidad de error, 417
tipos de, 346
virtudes de, 347
esquemas espectralmente eficientes, 347
estación base, 530
estadística suficiente, 321
estado cero, 481
estándar de codificación de audio MPEG-1, 9
adecuado para, 10
capacidades de, 234
desempeño de, 234
operación de, 235
estándar de codificación de audio MPEG-1, 9, 234, 237
estándar de codificación de imágenes JPEG, 8
estándar de codificación de video MPEG-1, 9
estándar de encriptado de datos (DES), 751–755, 759
estándar RS-232, 6–7
estéreo FM
especificación de los estándares, 124
multiplexado, 124–126
transmisión, 124
estimación de la lista de llaves, 753, 754
estimación de máxima verosimilitud de la fase de la portadora, 453–458
para resolver problemas, 449
estimación directa, 681
estimaciones sin ruido, 545
estrategia de paro y espera, 628
estructura armónica, 4
estructura de combinación de diversidad lineal, 545
estructura melódica, 4
evento elemental, 704
evento imposible, 704
evento nulo, 704
evolución de, 28–29
excitación cuantizada, 552
excitación de pulso ordinario, 552
expansión en series de Fourier, 317
expansor, 203
experimento aleatorio
descripción de, 703
características de, 703
extensión del umbral, 154
extracción, 261
extremo principal, 17

F

facsímil (fax)
en un modo de operación receptor, 420
objetivo de, 6
principio básico de, 6
factor de amortiguamiento relativo, 160
factor de ancho de banda de exceso, 441
factor de caída, 265
respuesta en el tiempo para, 265–266
respuesta en frecuencia para, 265–266
factor de escalamiento dependiente del sistema, 132–133
factores de eficiencia de la radiación, 520
Farnsworth, Philo T., 27
fase portadora, 403
fasor variable en el tiempo, 728
fasores, 532
FDM. *Vea* multiplexado por división de frecuencia
FDMA. *Vea* acceso múltiple por división de frecuencia
FEC. *Vea* corrección de errores directa
fenómeno de enmascaramiento auditivo, 234
fenómeno de trayectoria múltiple
en un ambiente de radio móvil, 18
formas de, 532–533
naturaleza de, 532

Fessenden, Reginald, 27
fibra óptica
como un medio de transmisión, 17
componentes de, 17
propiedades de, 17
ventajas de, 17
figura de mérito, 134
definición de, 132, 193
para la modulación de amplitud, 136
para la modulación de frecuencia, 147
filtrado, 128, 208–209
filtrado adaptable, 297
filtro acoplado, 248–252, 286
filtro acoplado no coherente, 406
filtro antitraslape, 187
filtro bidimensional acoplado, 378–379
filtro de banda estrecha ideal, 45
filtro de banda lateral residual (VSB)
respuesta en frecuencia de, 102
respuesta en magnitud de, 100–101
filtro de conversión binaria doblemente modificada, 273
filtro de conversión duobinaria, 268–269
filtro de cuadratura óptima, 378–379
filtro de formación de banda lateral residual (VSB), 102
filtro de formación de pulsos
propiedades deseables de, 396–397
respuesta al impulso gaussiano de, 398
filtro de lazo, 157, 159–160
filtro de predicción, 228
filtro de rastreo, 154
filtro de recepción, 259
filtro de reconstrucción, 187–188
filtro de respuesta al impulso de duración finita (FIR), 379
filtro de síntesis, 552
como parte del codificador, 551
componentes de, 552
filtro digital, segundo orden, 454–455
filtro en fase óptima, 378–379
filtro en fase pasobanda, 375, 378
filtro gaussiano, 397
filtro lineal invariante con el tiempo
como un filtro acoplado, 250–251
definición, 250–251
respuesta al impulso de, 44
uso de, 248
filtro óptimo
componente de, 378–379
respuesta al impulso de, 250–251
filtro pasobajas de premodulación, 396–397
filtro pasobanda, 98–99, 515
filtro posdetección, 143
filtro transmisor, 259
filtros lineales de preacentuación y desacentuación
aplicaciones en, 157
uso de, 157
firma digital
para sistemas de correo electrónico, 758–759
propiedades de, 759
uso de, 758–759
Fleming, John Ambrose, 27
flip-flops, 646
fluctuaciones espontáneas, 58–61
flujo de llaves, 744, 745
forma de onda, 21, 276
de códigos de línea importantes, 204–207
en modulación, 490, 492
forma de onda de señalización pasobanda, 729
forma de onda DS/BPSK, 490, 492
forma del pulso transmitido, 261
formación de pulsos, 247
formación espectral, 348
formas de onda de perturbación, 488
formato de modulación, 101–102
fórmula de Euler, 453
fórmula de Friis, 526
fórmula de interpolación, 186
fórmula de la suma de Poisson, 718
fotocátodo, 4
frecuencia de exploración de líneas, 5–6
frecuencia de la portadora, 92
frecuencia de Nyquist, 186–187, 262
frecuencia fundamental, 717
frecuencia instantánea, 110
definición de, 163
ecuación para, 108
frecuencia intermedia (FI), 128
frecuencia natural, 160
frecuencias de circuito telefónico, 3
frontera de capacidad, 602
FSK binario coherente convencional
con decodificación de 1 bit, 417
expresión de tasa de error de bits para, 417
FSK de Sunde, 381, 385–386, 388
FSK. *Vea* corrimiento de frecuencia por llaveo
fuente continua, 615
fuente del mensaje, 348
fuente discreta, 615
fuente discreta sin memoria, 570
extensión de, 572
propiedades de, 568
fuente extendida, 572
función argumento, 54
función básica, 451
función característica, 713
función de autocorrelación, 36, 43, 482
definición de, 35
evaluación de, 51
importancia de, 37
propiedades de la, 36–37
resumen gráfico de, 75–76
función de autocorrelación promediada en el tiempo, 42, 51
función de autocorrelación temporal, 284
función de autocovarianza, 36
función de Bessel, 735–739
comportamiento de, 114–115
contra el índice de modulación, 114
propiedades de, 735–737
función de Bessel modificada, 737–739
función de correlación de espaciamiento de frecuencia y espaciamiento de tiempo, 538
función de costo, 558
función de covarianza, 35–36
función de densidad de probabilidad condicional, 320–321, 383, 451–552, 710
función de densidad de probabilidad, 67–68, 255, 594, 708–709, 710
función de densidad de probabilidad conjunta, 594, 709–710
función de dispersión, 539
función de distorsión de la velocidad, 612–613, 616
aplicación de, 616
definición de, 613
función de distribución
de un proceso aleatorio estacionario, 34
propiedades de, 708
función de distribución acumulativa, 708
función de distribución conjunta, 33, 709

función de error complementaria, 255, 256, 334
función de forma del símbolo
definición, 353
densidad espectral de la energía de, 386, 395
función de formación de pulsos, 373
función de interpolación, 186, 427
función de probabilidad logarítmica, 452
definición, 322
definida para canal AWGN, 325
relación de, 322
función de transferencia de distancia, 666
función de transferencia variante con el tiempo, 536
función de una vía de puerta trasera, 755
función de una vía de transpuerta RSA, 758
función de verosimilitud condicional, 403–404
función delta de Dirac. *Vea* función delta
función delta, 62
propiedad de, 716–717
propiedad de filtrado de, 190, 320
función gaussiana, 397
función heterodina, 128
función hipergeométrica confluente, 740–741
función impulso, 62
función lineal, 54
función logarítmica, 322
función rectangular, 262
función senc, 262
función signo, 724
funcional
contra función, 54
definición, 54
funciones básicas pasobanda
propiedades de, 434–436
variaciones en tiempo de, 373
funciones de base ortonormal, 311, 315, 494
como una propiedad deseable, 440–441
desventajas de, 441
para MSK, 390
funciones de correlación cruzada, 40, 52
funciones de correlación, 35–41
funciones de la media, 35–39
funciones de muestreo, 32–33
funciones de verosimilitud, 322
funda de cloruro de polivinilo (PVC), 16

G

ganancia, 720
ganancia de codificación, 425
ganancia de codificación asintótica, 668, 673
ganancia de potencia, 523
concepto de, 521
de una antena, 520
definición de, 520
ganancia de procesamiento (GP)
definición, 229, 497
producción, 229
ganancia direccional, 520
generador de errores de fase, 459
generador de excitación, 551, 552
geometría celular hexagonal, 531
glotis, 4
GMSK. *Vea* corrimiento mínimo por llaveo de filtrado gaussiano
gráfica del flujo de la señal, 665–666
gráficas bipartidas, 685
GSM. *Vea* Sistema global para comunicaciones móviles
guías de coconjunto, 638–639

H

haces dispersados, 71
haz electrónico, 4
Hertz, Heinrich, 26–27
Hockham, G.A., 29
hoja de balance del presupuesto del enlace, 517
host, 13

I

imagen de televisión, 5–6
imagen estática, 4
imágenes
como fuente de información, 4
percepción de, 4
y el sistema visual humano, 4
imperfecciones del canal, 2
incremento del ruido, 283
índice de desempeño, 224
índice de modulación
definición, 110
restricción de, 112
valores de, 117–118
valores pequeños de, 119
inestabilidad, 208
información, 2
información binaria de señalización, 345
información extrínseca, 679
información intrínseca, 679
información lateral, 230
información mutua, 584–585
definición, 596
en el modelo de Shannon, 746
para ensamble continuo, 593–597
propiedades de, 585–587, 596
información redundante, 227, 575
instalación de transmisión digital, 215
integración
efectos benéficos de, 221–222
como una operación lineal, 223
integral de convolución, 42–44, 718
intensidad de la radiación, 519
intensidad media de la señal, 530
intercalamiento
definición de, 674
tipos de, 674
uso de, 675
intercalamientos irregulares, 691–692
interfaz capa a capa, 13
interfaz, 11
interferencia
como no intencional o intencional, 479
efecto de, 490
intensidad de, 148–149
potencia promedio de, 495
y desvanecimiento, 71–72
interferencia aditiva de código modulado, 493
interferencia aleatoria, 31–32
interferencia de acceso múltiple (MAI), 548
interferencia de intersímbolo (ISI), 259–261, 398
bajo el control del diseñador, 268
como deterioro del canal, 379
como un deterioro dominante, 279
como un efecto indeseable, 267
como una forma de interferencia, 296
condición de, 282–283
efectos de, 294, 296
efectos de minimización de, 260
en distorsión máxima, 288
superación de los efectos de, 441
y error de temporización, 266
y errores de bit, 247
y presencia de ruido, 294
interferencia de la imagen, 129

interferencia electromagnética (EMI), 17
Internet, 13–14
interpretación del llenado de agua, 610–611
interruptor digital, 215, 446
intervalo de Nyquist, 186–187
intervalo de seguridad, 441
intervalo de señalización, 568
invarianza, 331
invarianza rotacional de 90°, 422–423
inversión directa de matrices, 558, 559

J
jerarquía digital, 214
jerarquía TDM digital en Estados Unidos, 214–215
Joint Photographic Experts Group (JPEG), 8

K
Kao, K.C., 29
Kotel'nikov, V.A., 27

L
láser, 29
lazo de Costas
  generalización de, 454
  para recuperación de fase, 454
lazo de Costas recursivo, 454
  comportamiento convergente de, 461, 462
  comportamiento de adquisición de fase de, 459
  operaciones de, 458
  para sincronización de fase, 454
lazo local, 420
ley A
  capacidades de, 203
  definición, 202, 203
ley al cuadrado, 193
ley de compresión y expansión de 15 segmentos, 426
ley del inverso cuadrático, 519
ley exponencial, 193
ley m, 202–203
leyes de compresión, 202–203
leyes de expansión, 203
libro de códigos, 580
limitador de amplitud, 129, 142, 143
límite de Shannon, 602
línea digital de suscriptor (DSL)
  ambiente operativo de, 277, 447
  códigos de línea para, 280–281
  como aplicación creciente, 277
  contra módems de banda de voz, 446
  y pares trenzados, 277, 297
líneas de suscriptores digitales de velocidad muy alta (VDSL)
  uso de, 446
  ventajas de, 446
llave secreta
  contra llave pública, 759
  selección de, 745
llaves, 756
lóbulo principal, 368, 720
Lodge, Oliver, 27
lógica de corrección de módulo 2p, 364
longitud de palabra de código, promedio, 574
LPC de código excitado, 552
LPC excitado de pulsos múltiples, 551–552
LPC. *Vea* codificación predictiva lineal
Lucky, Robert, 28
luz, 6

M
malla de fase encadenada, 121
  complejidad de, 159–160
  componentes de, 157–160
  entendimiento de, 157, 158
  filtro de lazo en, 158
  forma más simple de, 159–160
  limitaciones de, 159–160
  modelo de, 158–160
  uso de, 157–160
mapeo de frecuencia en el tiempo, 9
mapeo de muchos a uno, 8
mapeo inverso, 589
máquina de estado finito, 654
Marconi, Guglielmo, 27
margen de perturbación, 499
margen de ruido, 581
matriz circulante, 442, 443
matriz de correlación, 40
matriz de identidad, 329
matriz de transición, 582
matriz de verificación de paridad, 634
matriz del canal, 582
matriz diagonal, 443
matriz ortonormal, 443
matriz unitaria, 443
matrizador
  generación de señal de diferencia, 125
  generación de señal de suma, 125
Maxwell, James Clerk, 26
mecanismo de audición, 4
media, 35
media geométrica, 436
medida de la distorsión, 199
medida de la distorsión de una sola letra, 612
mensaje, 155
método de discriminación de frecuencia
  etapas del, 98–99
  uso de, 100
método de la frecuencia relativa, 703–704
método del descenso más pronunciado, 225–226
mezclador
  componentes de, 103–104
  función de, 129
  operación de, 105
microceldas, 531
micrófono, 15–16
Middleton, D., 27
minimización de errores, 551
modelación psicoacústica, 9
modelo de criptografía de Shannon
  método de confusión, 749
  método de difusión, 749
  métodos de diseño, 749
modelo de propagación en el espacio libre, 518–523
modelo de receptor, 130
modelo de referencia de interconexión de sistemas abiertos (OSI), 11
modelo de ruido blanco gaussiano aditivo (AWGN), 516
modelo de transmisión pasobanda, 348–349
modelo del canal, 130
modelo del receptor de ruido, 130
modelo gaussiano, 55
modelo OSI, 11–13
modelos matemáticos
  clases de, 31
  en términos probabilísticos, 31
módem, 7
  como un dispositivo de conversión, 420
  configuración de, 421
  diseño de, 421
  e Internet, 420
  partes de, 420
módem analógico
  desempeño en ruido de, 429
  filosofía de diseño de, 429–430

operación limitada de, 429
módem asimétrico
configuración del, 425–426
diseño del, 426
módem digital
bases teóricas para el diseño de, 429
bidireccional, 428
capacidades de, 428
esquema de señalización para, 427
filosofía del diseño fundamental, 426
restricciones de diseño de, 426
solución a problemas de diseño, 428
una realización de, 426–427
y velocidades de datos, 429
módem en la banda de voz, 420
ambiente operacional de, 447
contra líneas de suscriptor digital, 446
módem V.32 estándar y esquemas de modulación alternativos, 421
módem V.32 y codificación no redundante, 423–424
cambios de fase en, 422–423
conmutación al modo QPSK, 424
y codificación de enramado, 423–424, 425
módem V.90, 431
módem, facsímil, 420
modo de decisión dirigida, 290–291
modo de entrenamiento, 290
modo de transferencia asíncrona (ATM), 14–15
modo de transmisión, 2–3
modulación, 19
etapas de, 490
operaciones de, 628
modulación angular, 20, 89–90
característica importante de, 107
clasificación de, 163
formas de, 108–109
provisiones de, 107
modulación codificada por enramado, 668–669
modulación de amplitud (AM), 20, 89–90, 422, 729
definición de, 90
figuras de mérito, 162–163
limitaciones de, 92–93
tipos de, 162–163
virtudes de, 92–93
modulación de amplitud completa, 162–163
modulación de amplitud en cuadratura (QAM), 97–98
constelación de cruce, 371
contra CAP, 369
modulación de amplitud en cuadratura M-aria (QAM), 369–373
modulación de amplitud residual, 112
modulación de amplitud/fase sin portadora (CAP), 373, 431
ancho de banda para, 375
idea detrás de, 369
modulación de banda lateral residual (VSB), 100–101, 163
modulación de banda lateral residual
definición de, 93
y su papel en la transmisión de la TV comercial, 101–102
modulación de banda lateral única (SSB), 98–100, 163
definición de, 93
en el multiplexado de división de frecuencia, 106
operación básica en, 103
modulación de corrimiento de amplitud por llaveo (ASK), 345
modulación de doble banda lateral-portadora suprimida (DSB-SC), 133, 134
definición de, 93
generada por, 94
transmisión de bandas laterales, 163
modulación de espectro disperso
comunicaciones seguras de, 479, 480
definición de, 479–480
para aplicaciones militares, 480
para proporcionar rechazo de trayectoria múltiple, 480
principios de, 480
técnicas de señalización conocidas como, 479
modulación de fase (PM), 20, 108, 729
modulación de fase diferencial, 422
modulación de frecuencia (FM), 20, 27, 729
capacidad de, 165
característica de, 149
casos de, 111
como un proceso no lineal, 109
definición de, 108–109
directa, 120–121
teoría de, 126
y mezcla, 500
modulación de frecuencia binaria, 397
modulación de frecuencia de banda ancha, 113–115
modulación de frecuencia de banda angosta, 111–113
modulación de frecuencia directa, 120–121, 396
modulación de frecuencia indirecta, 120–121
modulación de onda continua (CW), 20
efectos en la recepción, 130
familias de, 88–90
principios de, 162
técnicas para, 130
modulación de SSB, 134
modulación de un solo tono
y una señal de FM de banda ancha, 110
y una señal de FM de una banda angosta, 110
modulación delta (DM), 218–221
principios de, 218–219
simplicidad de, 223
ventaja de, 219, 237
y error de cuantización de, 220
y modulación de pulso, 237
y transmisores, 228
modulación delta adaptable (ADM), 232–235
modulación delta-sigma, 221–223
modulación diferencial adaptable por codificación de pulsos (ADPCM), 229–232
descripción de, 230–231
para señales de voz, 232
uso de, 231
modulación diferencial por codificación de pulsos (DPCM), 227–229
comparación del sistema, 228
idea básica de, 227
y modulación de pulsos digitales, 237
y transmisores, 228
modulación digital por pulsos
característica de, 236
forma básica de, 193
transmisión de, 183
modulación en amplitud por cuadratura (QAM), M-aria coherente, 464
modulación en cuadratura, 670

modulación espectralmente eficiente, 347
modulación lineal
definición de, 93
ejemplo de, 163
formas de, 93–94
tipos de, 93
modulación multicanal, 440–441
forma de, 431, 465
idea básica de, 431
modulación ortogonal no coherente, 407–409
caso especial de, 414
desempeño del ruido de, 407
receptor para, 408
modulación pasobanda, 729
modulación por amplitud de pulsos (PAM), 20, 188–191, 236
definición de, 188
y diseño del modulador, 277
y muestreo natural, 188
modulación por amplitud de pulsos discreta (PAM), 259
modulación por ancho de pulsos, 191
modulación por codificación de pulsos (PCM), 193, 615
como el método preferido, 20
como una forma de modulación de pulsos digital, 237
condición básica de, 194
costo de las ventajas, 217–218
definición de, 201
desempeño de, 227
para la codificación del habla, 229–230
requerimiento del ancho de banda, 218
uso de, 217–218, 560
ventajas de, 217, 237
modulación por duración de pulsos (PDM), 20, 191–192, 236
modulación por posición de pulsos (PPM), 20, 192, 236–237
modulación por pulsos analógica, 20
característica de, 236
forma óptima de, 193
transmisión de, 183
variaciones de, 236
modulación por pulsos discreta, 259
modulación por pulsos
como técnicas de codificación de fuente, 237
familias de, 183
forma digital estándar, 20
formas de, 191–193, 237
método utilizado para transmitir, 211
naturaleza de pérdidas de, 237–238
tipos de, 20
modulación senoidal, 112–113
modulación sigma-delta, 222
modulador de corrimiento mínimo por llaveo de filtrado gaussiano (GMSK)
e interferencia de intersímbolo, 398
pulso de formación de frecuencia de, 397
modulador de fase de banda angosta, 120
modulador de producto, 94, 98–99, 111, 490
modulador de pulsos, 211
modulador delta lineal, 221, 232–233
modulador, demodulador, 420
moduladores de amplitud en cuadratura, 433–434
momentos, 712
momentos centrales, 712
momentos de unión, 713–714
Morse, Samuel, 26
Motion Picture Experts Group (MPEG), 9, 234
movilidad, 18
movimiento ondulatorio, 18
MSK coherente
expresiones para, 417
tasa de error de bits para, 394
MSK de filtrado gaussiano, 396–400
muestras de cresta plana, 191
muestreo, 201–202
muestreo espacial, 4–5
muestreo instantáneo, 184
multiplexado, 105
de señales digitales, 214
definición de, 20
tipos de, 20–21
multiplexado de compresión de tiempo (TC)
esquema de, 278–279
uso de, 277–278
multiplexado de división de frecuencia ortogonal (OFDM), 447–448
aplicaciones de, 448
técnicas para transmisión, 344
uso de 447–448
multiplexado de portadora en cuadratura, 97–98, 354
multiplexado estéreo
como una forma de multiplexado por división de frecuencia, 124
en la transmisión de radio de FM, 124
multiplexado por división de código (CDM)
como método alternativo, 505
definición, 21
requerimiento de ancho de banda de, 505
multiplexado por división de frecuencia (FDM)
definición de, 20–21, 105
método de modulación en, 106
multiplexado por división de longitud de onda (WDM), 21
multiplexado por división de tiempo (TDM)
concepto de, 122
definición, 21,105
uso de, 211–212
multiplexor, 446
multiplexores de datos, 7
multiplexores digitales, 214–215
grupos principales de, 214
problemas de diseño, 215
multiplicador, 157, 646
multiplicador de frecuencia
componente de, 120
diagrama de, 120–121
multiplicadores de Lagrange
método de, 437
uso de, 609
multitonos discretos (DMT), 431, 440–443, 444–446
aplicaciones de, 446
idea básica de, 441
uso de, 441
y modulación multicanal, 447–448
música
como una fuente de información, 4
estructuras de, 4

N

National Television System Committee (NTSC), 6
niveles de división, 277
niveles de reconstrucción, 194
niveles de representación, 194
no linealidad
efectos de, 126
formas básicas de, 126
la presencia de, 126
no linealidad de fase, 127
nodo de verificación, 685
nodos variables, 685
North, D.O., 27

Noyce, Robert, 28
Nyquist, Harry, 27

O
objetivos de diseño del sistema, 3
octeto, 7
onda de FM
  con índice de modulación reducido, 152–153
  requerimiento de ancho de banda, 118
onda de modulación por codificación de pulsos (PCM) binaria, 193
onda modulada de banda lateral residual (VSB)
  componente de cuadratura de, 102–103
  métodos de generación, 100
onda modulada en frecuencia, 413
onda portadora, 19–20
onda portadora senoidal
  definición, 90
  forma de onda de, 490, 492
onda seno más ruido de banda angosta, 69, 69–71
onda senoidal, 88
ondas de FM de banda angosta, 112–113
operación de acoplamiento de patrón, 615
operación de multiplexado-demultiplexado digital, 214
operador de esperanza estadística, 711
oscilador controlado por cristal, 120
oscilador controlado por número, (NCO), 458
oscilador controlado por voltaje (VCO), 152–153, 157–158
oscilador de una sola clave, 384
oscilador local, 97
oscilador, controlado por cristal, 120

P
palabra de código, 627
  duración de, 212
  en forma binaria, 574
  longitud promedio, 574
palabra de código de canal, 21
palabra de código fuente, 21
PAM. *Vea* modulación en amplitud por pulsos (PAM)
PAM M-aria de banda base transmisión, 275–277
para evaluar, 347
parámetro de frecuencia, 128
parámetro de ganancia de lazo, 159
parámetro del tamaño del escalón, 225–226
parámetros de corrimiento, 531
parámetros de filtros cuantizados, 552
parámetros de la señal, 403
parámetros del canal, 587
pares de transformadas de Hilbert, 376, 724
partición, 670
patrón de error, 635
patrón de intensidad de la radiación, 520
patrón de potencia del campo, 521
patrón del ojo
  como una herramienta experimental, 293
  definición de, 293
  e información de desempeño, 293
  interpretación de, 293
PCM. *Vea* modulación por codificación de pulsos
percepción, 4
pérdida de trayectoria, 522
pérdida en el espacio libre, 522
pérdida por inserción, 16
perfil de autocorrelación de trayectoria múltiple, 537
perfil de intensidad de trayectoria múltiple, 539
perforación, 676
periodicidad, 718
periodo de muestreo, 184
periodograma, 51
persistencia de la visión, 5
perturbación, 493
  estrategia de, 495
  formas de onda de, 508
  tipos de, 508
perturbación, de un tono, 508
perturbación por ruido, banda parcial, 508
perturbación por ruido, barrera, 508
perturbación, multitonos, 508
perturbación, ruido de barrera, 508
perturbación, ruido de pulso, 508
perturbador de ruido de banda parcial, 508
perturbador de ruido de barrera, 508
perturbador de ruido de pulso, 508
perturbador de tonos múltiples, 508
perturbador de un solo tono, 508
peso de Hamming, 637, 666
Pierce, John R., 29
plano en frecuencia, 541
polinomio de verificación de paridad
  definición, 644
  recíproco de, 645
polinomio del generador, 645
  de un código cíclico, 643
  definición de, 656
polinomio del mensaje, 643
polinomio del síndrome, 647
polinomio irreducible, 505
polinomio primitivo, 505
porcentaje de modulación, 90
portadora modulada por datos, 500
portadora no modulada, 108
portadora piloto, 99
portadora y sistemas de sincronización temporal, 449
potencia de ruido de salida media
  definición de, 139
  ecuación para, 141
potencia de ruido de salida, 523
potencia, disponible, 61
potencia del ruido, 3, 61
potencia del ruido, cálculo de la salida promedio de, 151
  determinación de, 145–147
potencia promedio de la señal, 3
potencia radiada efectiva referida a una fuente isotrópica (EIRP), 521
potencia transmitida
  definición de, 3
  recurso de comunicación fundamental, 92
PPM. *Vea* modulación por posición de pulsos
precodificación de Tomlinson-Harashima, 430
predicción, 9
predicción adaptable lineal, 225–227
predicción adaptable
  con estimación directa, 231
  con estimación inversa, 231
  desventajas de, 231
  esquemas para, 230
  métodos preferidos de, 231
predicción lineal, 223–227
predictor adaptable, 225
preenvolvente
  componentes en cuadratura de, 374
  definición, 725, 730
  determinación, 726
  propiedad básica, 732
prefijo cíclico, 441
presupuesto del enlace, 518
principio de análisis mediante síntesis, 551–552

principio de combinación de relación máxima, 550
principio de invarianza rotacional
establecido como, 330
ilustración de, 330–331
principio de invarianza traslacional
aplicación de, 331–332
establecido como, 331
principio de reciprocidad en antenas, 521
principio de superposición, 718
privacía, 745
probabilidad
axiomas de, 704–706
de error de bit, 384
de error de símbolo, 258, 328–329, 352
de error, 254, 409
de una decisión correcta, 357
propiedades básicas de, 705–706
probabilidad condicional de error, 255, 556
probabilidad condicional de error, 352
probabilidad condicional del error de símbolo, 333
probabilidad condicional, 706–707
probabilidad conjunta, 706, 707
probabilidad de error, 328–329, 497–499
cota de unión en el, 332–335
en un canal ruidoso, 589
invarianza de, 329–331
probabilidad de error aparejado, 334
probabilidad de ocurrencia, 568
probabilidad de transición, 582, 616
probabilidad del error de símbolo
cálculo, 357
como una proporción, 358
contra tasa de errores de bit (BER), 335–336
definición de, 209, 310
evaluación de, 543
fórmula para, 256–257
para minimizar, 310, 346
simplificación de, 335
probabilidad del error de símbolo, 334
determinación de, 373
evaluación de, 346–347
fórmula par, 401
para la constelación de la señal, 337
probabilidad, condicional, 706–707
probabilidades *a priori*, 323–324, 583
probabilidades de estado, 681
problema de cercanía-lejanía, 548, 549
problema de cuantización óptimo, 199–201
problema de detección de señales, 322
enunciado como, 323
función de probabilidad para, 405
problema de la detección binaria, 403
problema de optimización restringido
definición de, 437
solución de, 609
problema de optimización, 438–440
problema de sincronización
métodos para resolverlo, 449
solución para, 493
problemas de carga, 438
procedimiento de estimación, 517
procedimiento de intercalamiento bit por bit, 214
procedimiento de optimización de lazo cerrado, 551
procedimiento de ortogonalización de Gram-Schmidt, 315–317
procedimiento de subespacio de dos pasos, 556
procesador de señales de arreglo lineal
para el diseño, 554
para el receptor, 554
requerimientos de, 554
procesador de señales del arreglo, 553, 556
procesador espacial, 534
procesador espacio–tiempo de la banda base, 556
procesador espacio-tiempo, 557
procesamiento espacial adaptable, 557
proceso aleatorio
a través de un filtro lineal
clases de, 75
definición de, 33
definición matemática de, 32–33
en sistemas lineales, 42–44
expectativas de, 41
invariante con el tiempo, 42–44
parámetro de, 75
potencia promedio de, 610
promedios de ensamble de, 41
propiedades de, 32
proceso de banda estrecha, 65–66
proceso de carga, 438
proceso de codificación
operaciones de, 9
pasos de, 551
uso de, 203–204
proceso de codificación de Huffman, 578, 579
proceso de comunicación, 1–3
proceso de comunicación del habla, 4
proceso de cuantización, 193–195, 236
en la generación de una onda PCM binaria, 193
ilustración de, 195
naturaleza no lineal, 198
resultados de, 20
proceso de decodificación
métodos de, 693
requerimientos de, 261
y generación de pulsos, 208
proceso de Markov, 678
proceso de modulación, 19–21
clasificación de, 20
definición de, 88
proceso de modulación de pulsos
pérdida de información y control del diseñador, 238
pérdidas de información en las que se incurre en el, 238
proceso de modulación híbrida, 374
proceso de modulación no lineal, 163
proceso de producción del habla, 4
proceso de ruido blanco gaussiano de media cero, 310
proceso de ruido blanco, 62
proceso ergódico, 41–42, 51
proceso estacionario, 33–34
contra estrictamente estacionario, 33
distintos nombres para, 36
proceso estocástico, 32
proceso estrictamente estacionario, 35
proceso gaussiano, 54–58
definición de, 57
en el estudio de las comunicaciones, 55
justificación matemática, 55–56
propiedades útiles de, 56–58
virtudes de, 55
procesos de iguales, 13
procesos de muestreo, 236
en la generación de una onda PCM binaria, 193
uso de, 184
y procesamiento de señal digital y comunicaciones digitales, 184
y sistemas de modulación por pulsos, 236
producto ancho de banda-duración, 721–722

producto cartesiano, 369–370
producto interior, 313, 314
producto tiempo-ancho de banda, 721-722
como un parámetro de diseño, 397
elección de, 398
producto vectorial, 681
promedio estadístico, 711–714
promedios totales
estimación de, 41
función de autocorrelación, 284
parámetro, 75
sustitución de promedios de tiempo para, 41
propagación, 4
propagación de errores
eliminación de la posibilidad, 273
fenómeno, 270
propiedad, 745
propagación guiada
canales basados en, 15
tipos de, 15
propagación libre
canales basados en, 15
tipos de, 15
propagación por radio
en áreas urbanas, 532–533
en el espacio libre, 512
propiedad cíclica, 442
propiedad de escalamiento en el tiempo, 722
propiedad de linealidad, 642
propiedad de Toeplitz, 225
propiedades compartidas de, 693
amplificador de bajo ruido, 515
análisis estadístico de, 689
palabras de código de bajo peso, 684
pasos de, 690–691
uso de, 684
proporción asimétrica, 282
proporción de pérdida de celdas, 14
protocolo de Internet (IP), 13–14
protocolo
de Internet, 13–14
tipos de, 13–15
protocolos para, 13–14
proveedor de servicios de Internet (ISP), 420
y comunicación entre PSTN, 425
y módems de voz, 420–422
y red pública de conmutación telefónica (PSTN), 420–425
prueba de hipótesis binaria, 405
PSK M-aria
comparación de, 419
características espectrales y de ancho de banda similares, 419
caso especial, 365
constelación de señales de, 365, 420
duración del símbolo de, 367
ecuación del error de símbolo para, 365–366
función de verosimilitud para, 452
requerimientos de ancho de banda de potencia para, 419–420
PSK M-aria coherente, 367
PSK. *Vea* corrimiento de fase por llaveo
PSTN digital, 426
pulso, 5
pulso binario codificado, 193
pulso coseno de semiciclo, 389
pulso de anulación, 5
pulso de banda base, 374
pulso en cuadratura pasobanda, 374
pulso en fase pasobanda, 375
pulso pasobanda, 375
pulso seno de semiciclo, 389–390
pulsos binarios codificados, 193
pulsos de sincronización, 5, 212
punto de escaneo, 5
punto de la señal recibida, 323
punto de la señal transmitida, 322–323
punto del mensaje, 322–323
puntos de muestreo, 32, 704

Q

QAM M-aria
desempeño de, 420
detección para, 371
energía transmitida en, 371
funciones en, 369
probabilidad del error de símbolo, 371
QAM. *Vea* modulación de amplitud en cuadratura
QPSK. *Vea* corrimiento de cuadrifase por llaveo
QPSK coherente
probabilidad del error de símbolo, 358
tasa de error de bits para, 417
QPSK, de compensación, 362
quadbits, 421
quanto, 194

R

radio celular
comunicaciones inalámbricas en el contexto de, 530
modelo idealizado de, 530
problemas de propagación de, 532
radio móvil, 18, 529
ráfaga de errores, 652
ráfagas, 7
ranura de tiempo, 513
rastreo, 493
razón de la desviación, 387
contra índice de modulación, 119
definición de, 119
razón de multiplicación de frecuencia, 121
razón SNR. *Vea* codificación de la decisión flexible de la relación señal a ruido, 630
realimentación de decisión, 270
receptor
como un detector óptimo de máxima verosimilitud, 436
de bobina móvil, 15–16
de un sistema de comunicación analógico, 88–89
desacentuación en, 154–155
desempeño en cuanto al ruido de, 387
modelo, 130
suposiciones de, 403
y preprocesamiento de la señal recibida, 64
receptor adaptable, 287
receptor bidimensional óptimo, 378
receptor CAP óptimo, 379
receptor cuadrático
ecuaciones de, 405
formas de, 405
receptor cuadrático óptimo, 403–405
receptor de AM
comparación de, 136–137
desempeño de, 136–137
detección de envolvente, 135–137
modelo de, 135
receptor de amplitud/fase sin portadora (CAP)
desempeño mejorado de, 379
en un ambiente desconocido, 379
implementación digital de, 379–380
receptor de bobina móvil, 15–16
receptor de correlación, 326–328, 329
receptor de corrimiento de cuadrifase por llaveo (QPSK), 360

receptor de Costas
componentes de, 96–97
control de fase en, 97
uso de, 96–97
receptor de doble banda lateral-portadora suprimida (DSB-SC)
comparada con un receptor de radio AM, 136–137
modelo de, 132–133
receptor de filtro acoplado
correlación, 326–327
parte detectora de, 328
receptor de modulación por codificación de pulsos (PCM), 258
receptor en cuadratura
canales de, 408
utilizando correlatores, 405–506
utilizando filtros acoplados, 405–406
receptor FMFB, 154
análisis de ruido de, 146
efecto de umbral, 152
modelo de, 142–143
punto de falla de, 149
receptor de FM
ruido en, 142
supresión de interferencia en, 148–149
receptor lineal óptimo, 282–287
interpretación de, 286
y transmisor, 286
receptor lineal, 248
desempeño de, 132
diseño de, 283
utilizando detección coherente, 132
receptor MSK, 394, 395
receptor no coherente, 405–406, 409
receptor óptimo
como un receptor de correlación, 326–327
diseño de, 310
parte detectora de, 328
subsistemas de, 326–327
receptor RAKE, 549–550
como un receptor de diversidad, 549
componentes de, 549
idea básica de, 549
técnicas de, 549
receptor superheterodino, 27, 128–129
componentes de, 128
diferencias entre AM y FM, 129
receptores de color
en el ojo humano, 6
tipos de, 6
recuperación de fase, 450
recuperación de fase portadora, 448, 458, 459–463
recuperación de la temporización del símbolo, 463–464, 465
recuperación de reloj, 448
recursos de comunicación, 3
recursos de red, 11, 14
red, 10
red de circuitos de conmutación, 11
controlada por, 11
establecimiento de conexión, 11
red de datos, 11
red de mapeo de tiempo a frecuencia, 235
red de pendiente, 143
red de televisión, 14
red digital de servicios integrados de banda ancha (B-ISDN), 14–15
calidad de servicio, 14
celdas en, 15
y red telefónica, 14
red interconectada, 13
red óptica síncrona (SONET), 15
red pública de conmutación telefónica (PSTN), 237, 420
como una red analógica, 420, 421
distorsión en, 286–287
uso eficiente de, 425
red telefónica, 27
como una red de comunicación, 10
primer servicio comercial, 28
propósito fundamental, 15
redes de banda ancha, 14–15
redes de comunicación, 10–15
redundancia
formas básicas de, 9
suma de, 626
uso controlado de, 628
redundancia de codificación entrópica, 9
redundancia de interpixeles, 9
redundancia de intertrama, 9
redundancia psicovisual, 9
Reeves, Alec, 27
regeneración de la señal, 208
regeneración, 208
región de integración, 332
registro de corrimiento realimentado, 480, 481
registro de corrimiento, 481
regla de Carson
para evaluación aproximada, 163
y la curva universal, 119
regla de decisión de máxima verosimilitud
para un canal AWGN, 325
propósito de, 325
regla de decisión, 350–351, 357
aplicación, 382
como la regla de máxima verosimilitud, 324
como la regla MAP, 323–324
definición, 661
usada por el detector coherente, 497
regla de decisión óptima, 323–324
regla de decodificación de máxima verosimilitud, 661
regla de decodificación, 660
regla de Leibniz, 257
regla de máxima verosimilitud, 324
regla de probabilidad *a posteriori* (MAP) máxima, 323–324
regularidad estadística, 703
relación de densidad espectral de la energía al ruido de la señal, 252
relación de entrada señal a ruido, 134
definición de, 131
ecuación para, 497
relación de la densidad de la energía del símbolo a la densidad espectral del ruido, 502
relación de la energía de bit a la densidad de ruido, 384
relación de la señal a ruido de banda base, 154
relación de señal a enmascaramiento (SMR), 234
relación geométrica señal a ruido, 436
relación portadora a ruido, 137–139, 144
baja contra alta, 149
contra relación señal a ruido, 151
definición, 138, 150
grande contra pequeña, 141–142
nivel de operación, 138
relación ruido a enmascaramiento (NMR), 234
relación señal a ruido
de la fuente, 524
de un receptor de FMFB, 153
definiciones básicas de, 3, 130–132
en la salida del dispositivo, 524
limitación de, 261

relación señal a ruido de pulso máximo
de un filtro acoplado, 251
definición, 248–249, 250
relación señal a ruido de salida
cálculo de, 151
contra relación de portadora a ruido, 151
de receptor de AM, 136
de un cuantizador uniforme, 197
definición de, 131, 145
determinación, 133
ecuación para, 139, 497
factor de mejoramiento en, 155–156
incremento de, 155
para evaluar, 135
relación señal a ruido del canal, 134
definición de, 132
fórmula para, 147
para AM, 135
relaciones de Einstein-Wiener-Khintchine, 46
repetidor regenerativo, 208
replicación de la propiedad de la función delta, 717
representación en series de Fourier, 114
representación geométrica de señales, 311
representaciones del espacio de la señal de la señal de interferencia (perturbación), 493
de la señal transmitida, 493
requerimiento de la media condicional, 200
resistor ruidoso, 60
resolución, 6
resolución de la incertidumbre, 568
respuesta al impulso, 656, 718
respuesta al impulso del canal, 291
respuesta en el tiempo, 266
respuesta en fase, 719
respuesta en frecuencia
elección de, 155
para denotar, 44–45
respuesta en magnitud, 45, 608, 719
resultado importante de, 591
retardo de celda
definición, 14
variación del, 14
retardo de la envolvente, 16
retardo de la transmisión, 260
retardo de tiempo de la propagación, 516
retardo promedio, 540
retardo unitario, 219
retrazado horizontal, 5
retroceso. *Vea* retrazado horizontal
reuso de frecuencia, 530
ruido
ausencia de, 261
cálculos, 61
como señales indeseables, 58
definición de, 3
efecto de, 3
en sistemas de comunicaciones, 58
fuentes de, 3, 58
minimización de los efectos de, 157, 260
presencia de, 589
señales en, 322–326
ruido blanco, 61–63
características de, 61–62
densidad espectral de potencia de, 61–63
función de autocorrelación de, 61–62
propiedades matemáticas de, 62
ruido blanco gaussiano, 62
distribuido idénticamente, 334
proceso, 392
ruido blanco gaussiano aditivo (AWGN), 378
como deterioro del canal, 379
detección de señal en, 349
y diseño del receptor, 337
ruido blanco gaussiano limitado en banda, 608
ruido de banda angosta, 64
caracterización de, 64
componentes de, 64, 65–66
efectos de, 64
representación de, 64–66, 67–69
representación y sistema de coordenadas para, 67–68
ruido de cuantización, 195–197, 228
como función del tiempo, 195
control de diseñador de, 209
en modulación delta, 221
en sistemas PCM, 209
y percepción del oído humano, 9
ruido de cuantización, 229
ruido de disparo, 58–60
ruido del canal, 31–32
ausencia de, 228
condición de actuación aislada, 282–283
desvanecimiento del efecto de, 210
efectos de, 209, 296
en sistemas PCM, 209, 253
fuente de, 32
incertidumbre debida a, 403
y error de bit, 248
ruido granular, 220
contra ruido de cuantización, 221
y distorsión, 221
ruido térmico, 32, 60
ruido tipo impulso, 446
ruteador
definición, 13
en dispositivos host, 10
propósito principal de, 10

S

salida, 406
propiedades de, 251–252
salida de canal fija, 582
salida de tres niveles, 267–268
salida del canal, 72
salida del demodulador, 92
salida del discriminador, 145, 155
salto aleatorio, 500
salto de frecuencia rápido, 502–503, 504
salto de frecuencia, 500
satélite
en órbita geoestacionaria, 18
para comunicaciones, 29
servicios de, 18–19
satélite de transmisión directa (DBS)
satélite Telstar, 29
satélites de comunicación
"segunda generación", 515
en órbita geoestacionaria, 19
notas históricas, 26–29
papel de los, 19
satélites de transmisión directa (DBS)
uso de, 517
SDMA, 516
secuencia binaria aleatoria, 482
secuencia de datos binarios, 359
secuencia de longitud máxima
definición, 482
función de autocorrelación de, 482
propiedad de balance de, 482
propiedades de, 482–484
selección de una, 484
secuencia de pseudo ruido (PN), 480, 488
como una secuencia de entrenamiento, 288
componentes de, 288
generación y propiedades de, 480

secuencia ordenada
pseudoaleatoriamente, 500
secuencia PN
como una secuencia binaria independiente e idénticamente distribuida (iid), 496
como una señal de referencia, 550
propiedades de correlación de, 506
secuencias (código) de Gold
clases de, 505
propiedades de correlación de, 507
secuencias de longitud prima de Mersenne, 485
segmentos, 11
Vea también paquetes
segundo retroceso. *Vea* filtro digital de retroceso de segundo orden, 454–455
segundo teorema de Shannon, 616
seguridad de la transmisión, 490
seguridad perfecta, 746–747
selectividad de fase, 723–725
sensibilidad de amplitud, 90
sensibilidad de fase, 108
señal
definición de, 3–4
detección en ruido, 322–326
dimensiones de, 3–4
recibida contra transmitida, 2
señal (onda) de AM de banda angosta, 112–113
señal (onda) de FM de banda angosta, 111–112, 113
señal (onda) modulada, 95
señal (onda) moduladora senoidal, 110
señal (onda) modulante, 88
señal analítica. *Vea* preenvolvente
señal analógica portadora de información, 184
señal antípoda, 349
señal compuesta, 105, 149
señal de ajuste, 453
señal de AM
demodulación de, 162–163
y comparación de señal de FM, 164
señal de amplitud modulada, 89–90
señal de banda ancha, 7
señal de banda base, 88, 95
señal de corrimiento de frecuencia por llaveo (FSK), 386
señal de corrimiento de frecuencia por llaveo en fase continua (CPFSK), componentes de, 389
fase de, 388
relación de desviación de, 388
representación de, 387
señal de corrimiento mínimo por llaveo (MSK)
espectros de potencia de, 360
señal de corrimiento mínimo por llaveo de filtrado gaussiano (GMSK), 397
compactación espectral de, 398–400
espectro de potencia de, 400
señal de crominancia, 6
señal de error
cálculo de, 457
definición de, 288, 453
para recuperación de tiempo, 457
uso de, 557
señal de FH/MFSK transmitida, 500
señal de FM
análisis espectral de, 110
ancho de banda efectivo para, 117–119
características fundamentales de, 110
demodulación de, 121–124
detección de, 397
diferenciación de la señal de AM, 109
en la teoría y en la práctica, 117
envolvente compleja de, 114
espectro de, 115
frecuencias laterales de, 117
generación de, 120–121
potencia promedio de, 115
propiedades deseables, 397
señal de FM de banda ancha, 118
señal de FM de un solo tono, 112–113
señal de FM de un tono relativo, 112–113
señal de luminancia, 6
señal de manipulación de corrimiento de cuadrifase (QPSK)
constelaciones comúnmente utilizadas para, 362
espectro de potencia de, 360–361
filtrado, 361
fluctuaciones de amplitud, 362
observaciones de, 360
producción de interferencia, 396
transiciones de fase de, 361
y señal PSK binaria, 360
señal de modulación por codificación de pulsos (PCM), 218
señal de perturbación, 488
señal de referencia, 557
señal de salida del arreglo, 557
señal de salida media, 139
señal de TV transmitida, 102
señal de video, 4–5, 6
señal del habla, 4
como bipolar, 6
límites de, 16
señal del mensaje, 132–133
descripción de, 2
generación de, 2
señal dispersa de secuencia directa de corrimiento de fase por llaveo binaria (DS/BPSK), 490, 492
señal FSK binaria
con fase continua, 385–386
densidad espectral de potencia de banda base de la, 386
detección de, 387
espectros de potencia de, 353, 385–386
para generar, 384
señal FSK de Sunde, 386
señal FSK M-aria
análisis espectral de, 401, 402
eficiencia del ancho de banda de, 401–402
espectros de potencia de, 401
requerimientos del ancho de banda, 401
señales ortogonales de, 401–402
señal lenta de FH/MFSK, 501
señal limitada en banda
definición, 427
teorema de muestreo para, 186–187
señal M-aria, 402
señal modulada (onda) con doble banda lateral-portadora suprimida (DSB-SC), 95, 96
señal modulada angular, 387
forma de onda de, 89–90
interpretación de, 108
señal modulada de banda lateral única, 98–99
señal modulada de SSB, 99
señal modulada híbrida, 123
señal modulada por amplitud de pulso, 429
señal modulada por pulsos, 237
señal MSK, 391
características de, 396
demodulación de, 394
detección de, 394
espectros de potencia de, 394–396, 398

formas posibles de, 390
generación de, 394, 396
probabilidad de error, 392
propiedades de, 396
señal muestreada ideal, 184
señal multiplexada, 125–126
señal PAM
desempeño de, 191
forma de onda de, 188–189
generación de, 188–189
muestreo de, 189
transmisión de, 191
señal pasobanda
componentes de, 728, 730
representación de, 113, 726–729
transformada de Hilbert y, 731
señal PM, 109
señal portadora de información, 31–32, 88
en el dominio digital, 277
multiplicado por la señal PN, 488
señal PSK M-aria
densidad espectral de potencia de banda base de, 367
eficiencia del ancho de banda de, 368
eficiente espectralmente, 402
espectros de potencia de, 367
señal QAM M-aria, 372
señal QPSK desfasada p/4
demodulación de, 365
residente en uno de ocho posibles estados de fase, 362
señal recibida
componentes de, 31–32
valor medio de energía, 543
señal transmitida de banda ancha, 490
señal transmitida, 280–281
señales con fase desconocida, 403–406
señales de energía, 312
señales de energía mínima, 331–332
señales de televisión, 101–103
como señales de banda ancha, 7
formato de modulación de, 101–102
señales de voz, 99
señales digitales, 214
señales DQPSK desfasadas p/4, 364
señales musicales
contra señales de voz, 4
y ancho de banda del canal, 4
señales periódicas, 717–718
señales simples, 342
señales transortogonales. *Vea* señales simples
señalización activado-desactivado, 205
señalización binaria
esquema, 407
tasa de error de bits, 543–544
señalización bipolar con retorno a cero (BRZ), 207
señalización con inversión alternada de marca (AMI), 207
señalización con retorno a cero (RZ) unipolar
característica de, 207
desventajas de, 207
señalización de división de fase, 207
señalización de respuesta parcial, 269, 274–275
señalización polar sin retorno a cero (NRZ)
desventajas de, 205–206
sistema PCM binario basado en, 253
señalización sin retorno a cero (NRZ) unipolar, 205
serie de Fourier exponencial compleja, 717
"servicio del mejor esfuerzo", 14
servicio telefónico convencional (POTS), 281–282
Shannon, teorema de capacidad de Claude, 611
y la "teoría matemática de la comunicación", 27–28
y los fundamentos teóricos de las comunicaciones digitales, 27–28
Shockley, William, 28
símbolo, 2
símbolo de innovación, 581
símbolos de control
en ASCII, 6
para fines de comunicación, 6
para impresión de caracteres, 6
símbolos DQPSK desfasados p/4, 363
símbolos igualmente probables, 571
simplicidad y provisión de, 517
sincronización, 448–450, 493
como un problema de estimación de parámetros estadísticos, 449
de los relojes del transmisor y el receptor, 212
en un sistema TDM, 212
implementación de, 449
método algorítmico (moderno), 449
método clásico para, 449
métodos básicos de, 448
proceso de, 448
sincronización asistida por datos, 449
sincronización de portadora, 448
sincronización de retardo temprano-tarde recursivo, 463–464
sincronización de temporización, 455
sincronización del símbolo, 448
sincronización fina, 493
sincronización sin asistencia de datos, 449
sincronizador de retraso temprano-tarde sin asistencia de datos (NDA-ELD), 458
síndrome
cálculo del, 646–648
importancia del, 635
propiedades del, 635–636
sintetizador de ruido de banda angosta, 64–66
sintonización de la frecuencia portadora, 128
sistema auditivo, 235
sistema auditivo humano, 234
sistema CAP de alto desempeño, 375
sistema criptográfico
clases de, 744
clasificaciones de, 759
componentes de, 743
definición de, 743
servicios de, 742
sistema cuaternario
diagrama visual para, 294–295
salida de, 276
sistema de acceso múltiple
interferencia, 548
objetivo de, 549
sistema de amplitud/fase sin portadora (CAP)
aplicación de, 380
estructura básica de, 378–379
estructura de, 378–379
modulación, 369, 373, 375
receptor, 379–380
transmisor, 378
sistema de comunicación
análisis de ruido, 3, 64, 523
característica común de, 3
diseño de, 21
elementos de, 2
fuente de limitaciones, 248–252
propósito de, 19, 88
recursos primarios de, 3
transición de analógico a digital, 183

sistema de comunicación analógico
diseño de, 22
razones para el estudio de, 23
uso de, 21
sistema de comunicación digital binario, 403–405
sistema de comunicación inalámbrica
movilidad de, 559
problemas al utilizar MSK, 396
requerimientos prácticos de, 396
sistema de comunicación por radio, 31
sistema de comunicación por satélite geoestacionaria, 514–515
sistema de comunicaciones de espectro disperso
rechazo de interferencia, 479
requerimiento de, 493
ventaja de, 479
sistema de comunicaciones por satélite, 18
capacidades singulares del sistema, 18–19
cobertura global, 559
confiabilidad de, 512
diseño de, 517
sistema de demultiplexado, 125–126
sistema de distribución de clave pública de Diffie-Hellman, 756
sistema de espectro disperso de banda base, 488
sistema de FM
efectos no lineales en, 126–128
énfasis en, 154–156
Vea también sistema de modulación de frecuencia (FM)
sistema de llave pública, 759
sistema de llave secreta, 759
sistema de modulación de amplitud
análisis de ruido de, 135
proceso de, 90–91
sistema de modulación de frecuencia (FM)
análisis de ruido de, 142–147
similitudes con el sistema PPM, 193
sistema de modulación de onda continua (CW), 89–90
comparación de, 132
componentes de, 88–89
desempeño de, 164
ruido en, 130
sistema de modulación delta adaptable, 234–235
sistema de modulación por codificación de pulsos (PCM), 218
características de, 210
consideraciones de ruido en, 209–210
descripción de, 201–209
desempeño influenciado por el ruido, 209
e interferencia, 210
operación de, 212
operaciones básicas de, 201
sistema de modulación por posición de pulsos (PPM)
análisis de ruido de, 193
contra sistema de modulación de frecuencia, 193
desempeño de, 193
sistema de modulación, codificado en binario, 197
sistema de multiplexado de portadora en cuadratura, 98
sistema de probabilidad, 704
sistema de transmisión, 128
sistema de transmisión de datos binarios de la banda base, 259
sistema de transmisión de datos binarios, 285
sistema de transmisión de datos multicanal, 433, 434
sistema de transmisión de datos
asíncronos contra síncronos, 7
capacidades de, 446
desempeño de, 293
sistema de transmisión de pulsos de banda base y características fijas, 297 desempeño de, 296
fuente de errores de bit, 259
y relación señal a ruido, 297
sistema de transmisión digital pasobanda, 344
degradación del desempeño de, 544
evaluación del desempeño de, 335
sistema de transmisión multicanal, 437–438
sistema de transmisión óptico, 15
sistema DS/MPSK, 508
sistema FDM
diagrama de bloque de, 105–106
pasos de modulación en, 107
sistema FDMA, 516
sistema FH/MFSK
efecto de perturbación en el receptor, 502
error de símbolo en, 502
rápido contra lento, 502–503
sistema FSK binario coherente, 384
caracterizado por, 381
generación y detección de, 384
receptor, 384–385
sistema FSK binario, 381–386
Sistema global para comunicaciones móviles (GSM), 548
eficiencia de la trama de, 548
sistema de comunicación inalámbrica, 548
sistema ideal, 601
sistema lento de FH/MFSK, 502
sistema lineal, 718
sistema lineal invariante con el tiempo, 719–720
sistema M-ario, 276
sistema MSK, 391
sistema NTSC, 6
sistema PAM, 191
sistema PAM binario de banda base, 259
sistema pasobanda
análisis de, 730–734
respuesta al impulso, 731
sistema PSK binario coherente, 384
características de, 350
error de bit de, 357
receptor, 352
señales, 352
sistema QPSK coherente
especificaciones del, 458
señales de, 359
sistema realimentado de segundo orden, 160
sistema realimentado de segundo orden, 160
sistema Rivest-Shamir-Adleman (RSA), 757
sistema robusto, 22
sistemas CDM, 505
sistemas criptográficos de llave pública, 755
sistemas de comunicación inalámbricos digitales, 550–551
sistemas de modulación de pulsos, 236
sistemas de modulación lineal, 729
sistemas de televisión de antena comunitaria (CATV), 17
sistemas de televisión por cable, 17
sistemas de transmisión de datos pasobanda determinación de la eficiencia del ancho de banda de, 348
objetivo de, 346
sistemas multiplexados, 347

sistemas PAM M-arios
complejidad del diseño de, 277
consideración de, 276
en un ancho de banda de canal, 276
requerimientos de potencia de, 276–277
sistemas PSK M-arios, 449–450
sobremodulado, 90
solicitud de repetición automática (ARQ), 628–629
filosofía de, 628
para la detección de errores, 628
tipos de, 628–629
solicitud de repetición automática de paro y espera, 628
solución del llenado de agua, 438
SONET, 15
Strowger, A.B., 27
subredes, 13
subsistemas de receptor óptimo, 326–327
subtramas, 552
sumadores, 646
superheterodino, 128
suposición gaussiana, 498
supresión de interferencia, 148–149
supresión de señal débil, 142

T

tabla de cuantización, 8
tamaño del escalón, 194
tasa de código, 590, 627, 654
tasa de datos del canal, 627
tasa de desvanecimiento, 541
tasa de división, 501
tasa de error de bits (BER)
de esquemas de modulación digital, 417
en la regeneración de señales, 208
para la FSK binaria coherente, 384
probabilidad de, 23
suposiciones de, 209
y probabilidad del error de símbolo, 335–336
tasa de error, 253
tasa de muestreo, 184, 201
tasa de señalización, 276
TDMA. *Vea* acceso múltiple por división del tiempo
técnica duobinaria, 274–275
técnicas de acceso múltiple
característica común de, 514
definidas como, 513–514
ideas detrás de, 513, 514
y recursos de comunicación compartidos, 513
técnicas de combinación de receptor múltiple, 544
técnicas de diversidad del espacio, 546
técnicas de diversidad, 544–547
desempeño con, 546–547
técnicas especializadas, 559
técnicas de espectro disperso
como espectro disperso de secuencia directa, 490
contra técnicas de modulación estándar, 480
en la transmisión pasobanda, 490
técnicas de modulación digital M-aria, 419–420
técnicas de modulación digital
operación de, 448
tipos de, 345–346
técnicas de modulación estándar, 480
técnicas de preacentuación y desacentuación no lineales, 157
tecnología de circuito digital, 189
tecnología de conmutación de celdas, 14
telégrafo, 26
televisión analógica, 5
temperatura de ruido equivalente, 61, 524–525
temporización del símbolo, 455, 458, 463–464
teorema de Campbell, 60
teorema de codificación del canal, 589–593, 616, 630–631
aplicación de, 591–192
característica no satisfactoria de, 631
importancia de, 592
operaciones de, 628
segundo teorema de Shannon, 590–591
teorema de fuente de codificación, 574–575, 612, 616
en el primer teorema de Shannon, 574–575
longitud promedio de palabra de código de, 611
teorema de Gold, 505
teorema de irrelevancia, 321–322
teorema de la capacidad, 616
aplicación de, 607
argumento para, 599–600
como un canal de ruido coloreado, 607
implicaciones de, 601–603
interpretación del llenado de agua de, 610
parámetros del sistema de, 599
y canales gaussianos, 597
teorema de la capacidad de información de Shannon, 23–24, 433
teorema de la capacidad de Shannon, 611
teorema de la capacidad del canal, 599
teorema de la energía de Rayleigh
definición de, 251
uso de, 251–252
teorema de límite central, 60, 498
definición de, 56
y el proceso gaussiano, 55
teorema de máxima transferencia de potencia
aplicación del, 61
uso del, 61
teorema de muestreo, 201, 236
de un sistema de modulación de pulsos, 186–187
deducción de, 186–187
esencia de, 184
forma equivalente no uniforme recurrente de, 427
para señales de banda limitada, 186–187
teorema de potencia, 520
teorema de recursión, 682
teorema de separación, 681
teoría de distorsión de la velocidad, 611–613
aplicación de, 612
parámetros principales de, 613
y teoremas de codificación de Shannon, 612
teoría de la codificación, 28, 676
teoría de la información
límites fundamentales en, 567
teoría de la información de Shannon, 617
teoría de lazo de cierre de enlace, 157
teoría de los códigos para control de errores, 485
teoría de probabilidad, 703, 705
teoría del análisis espectral de procesos aleatorios, 46
tercer teorema de Shannon, 599
tercer teorema fundamental de Shannon, 616
término de diferencia de frecuencia, 158
término de doble frecuencia, 158

texto cifrado, 742
texto codificado, 6
texto simple, 742
tiempo de correlación inversa, 37
tiempo de excitación, 718
tiempo de memoria del sistema, 718
tiempo de respuesta, 718
tiempo de seguridad, 547
tiempo discreto, canal gaussiano sin memoria en, 597
tipo de forzamiento a cero, 556
tracto vocal, 4
trama(s), 552
  elaboración de, 5
  estructura, 547
  método de sincronización, 215–216
transformada coseno discreta (DCT), 8
transformada de Fourier
  de señales periódicas, 717–718
  definición de, 715
  inversa, 715
  propiedades de, 716
  teoría de, 716
transformada de Fourier en tiempo discreto, 185
transformada de Hilbert, 374, 408, 723–725
  de una señal, 724
  propiedades de, 725
transformada discreta de Fourier (DFT), 445
  definición, 442
  y procesamiento de señal digital, 443
transformada discreta inversa de Fourier (IDFT), 442, 445
transformada inversa de Fourier, 186, 715
transformador híbrido
  circuito simplificado de, 278–279
  definición de, 278
transiciones de la señal, 207
transistor, 28
transmisión asíncrona, 7
transmisión binaria sin distorsión en banda base, 261–262
transmisión de audio digital, 448
transmisión de datos digitales, 247
transmisión de datos directa, 281–282
transmisión de datos en sentido directo, 281–282
transmisión de datos pasobanda, 344
  aplicaciones de, 344
  canal de comunicación utilizado para, 344
  sobre canales no lineales, 345
  técnicas alternativas para, 465
transmisión de dos vías, 106
transmisión de información, 581
transmisión de pulso de banda base ideal, 262
transmisor
  a propósito de, 2
  de potencia limitada, 597
  de un sistema de comunicación analógico, 88–89
  localización de, 2
  mediante operaciones de combinación, 414
  uso de preacentuación en, 154–155
transmisor de amplitud/fase sin portadora (CAP), 378
transmisor de corrimiento de cuadrifase por llaveo (QPSK), 359
transmisor FSK binario, 384–385
transmisor MSK, 394, 395
transpondedor, 19, 514–515
transposición hermitiana, 443
transreceptor, 278
traslación de energía mínima, 332
traslación de frecuencia, 103–105
traslape, 187
trayectoria de la transmisión, 212
trayectoria múltiple
  fenómeno físico de, 513
  presencia de, 513
trayectorias sobrevivientes, 662

U

umbral de decisión, 194
umbral de enmascaramiento, 9, 234, 234–235
umbral de error, 209–210
umbral óptimo, 257
Ungerboeck, G., 28
unidad de empaquetado de tramas, 236
unión de eventos, 333

V

valores de umbral preestablecidos, 277
van Duuren, H.C.A., 28
Van Vleck, J.H., 27
variabilidad de la señal, 530
variable aleatoria, 33, 594, 708–710
  definición de, 708
  descripción de, 708
  desviación estándar de, 712
  distribución de, 55–56
  media de, 711
  valor esperado de, 711
  varianza de, 712
variable aleatoria continua, 594, 708
variable aleatoria de distribución gaussiana, 54
variable aleatoria discreta, 708
variable aleatoria gaussiana, 54, 58
varianza, 579
VDSL. *Vea* líneas de suscriptores digitales de velocidad muy alta
vector de código, 660
vector de código transmitido, 660
vector de error, 635
vector de información, 633
vector de la señal, 311
vector de observación, 321, 331
vector de peso, 554
vector de Poynting, 519
vector de ruido
  características estadísticas de, 330
  matriz de covarianza de, 330
  realización particular de, 323
vector de ruido rotado, 330
vector del mensaje, 660
vector del síndrome de error, 635–636
vector *N*-dimensional, 311
vector recibido, 660
vectores, 633
vectores aleatorios, 594
vectores de renglón, 633
velocidad angular, 108
velocidad angular constante, 108
velocidad de salto, 501
velocidad de señalización de datos, 426
velocidad del símbolo, 501
velocidad primaria, 214–215
ventana de decodificación, 663
video por demanda, 9, 282
VLSI. *Vea* circuitos integrados a muy grande escala de integración (VLSI)
von Neumann, John, 28
VSB. *Vea* banda lateral residual

W

World Wide Web, 28–29
  y características de, 421
  y trabajo pionero, 28

Z

Zworykin, Vladimir K., 27