

1. The first part of the paper is devoted to the study of the properties of the function $f(x)$ defined by the equation

$$f(x) = \int_0^x \frac{1}{1+t^2} dt$$
 for $x \in \mathbb{R}$. It is shown that $f(x)$ is an odd function and that it satisfies the inequality

$$f(x) \leq \frac{\pi}{2} \quad \text{for } x \geq 0.$$
 2. In the second part, we consider the function $g(x)$ defined by the equation

$$g(x) = \int_0^x \frac{t}{1+t^2} dt$$
 for $x \in \mathbb{R}$. It is shown that $g(x)$ is an even function and that it satisfies the inequality

$$g(x) \leq \frac{\pi}{4} \quad \text{for } x \geq 0.$$
 3. Finally, we study the function $h(x)$ defined by the equation

$$h(x) = \int_0^x \frac{t^2}{1+t^2} dt$$
 for $x \in \mathbb{R}$. It is shown that $h(x)$ is an odd function and that it satisfies the inequality

$$h(x) \leq \frac{\pi}{4} \quad \text{for } x \geq 0.$$

R621281
r 452 e
V. 3, p. 1

Enciclopedia de la **ELECTRONICA** **INGENIERIA Y TECNICA**

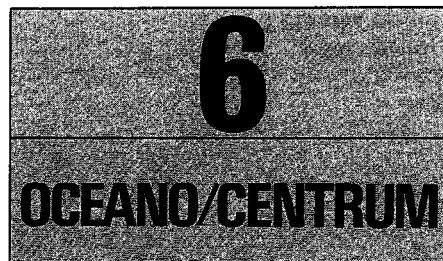
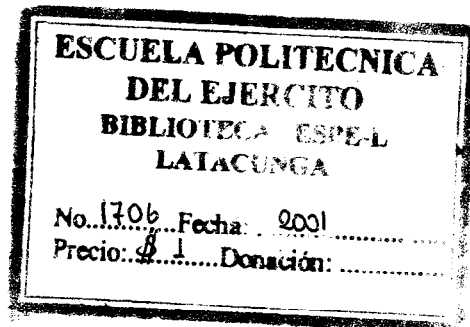
Charles Belove

**Departamento de Ingeniería Eléctrica y Computación
Florida Atlantic University
Boca Raton, Florida**

Director de la versión en español

Francisco Paniagua B., I.M.E.

Universidad Nacional Autónoma de México
Consultor Editorial de Ingeniería Eléctrica
Miembro de la U.S. Metric Association



Es una obra del
GRUPO EDITORIAL OCEANO

Presidente
José Lluís Monreal

Director General
José M.^a Martí

Director General de Publicaciones
Carlos Gispert



Versión española de la edición original de Charles Belove "Handbook of Modern Electronics and Electrical Engineering" publicada por John Wiley & Sons, Nueva York, EE.UU.

VERSIÓN ORIGINAL

Consejo Editorial: Phillip Hopkins
Lockheed Engineering and
Management Services, Co., Inc.
Houston, Texas

Edward Nelson
New York Institute of Technology
Old Westbury, Nueva York

Milton Rosenstein
New York Institute of Technology
Old Westbury, Nueva York

Stanley Shinnars
Jericho, Nueva York

VERSIÓN ESPAÑOLA

Supervisión Editorial: Pedro Basurto Samperio

Edición: Begoña Robles

Dirección Editorial: Marta Bueno

Traducción: Roberto Palacios Martínez
Licenciado en Ciencias
Universidad Autónoma de Baja
California

Hugo Villagómez Velázquez
Licenciado en Física y Matemáticas
Instituto Politécnico Nacional, México
Doctor en Ciencias
Universidad de París, Francia

Juan Carlos Vega Fagoaga
Ingeniería en Sistemas

José Rafael Blengio Pinto
Médico Cirujano
Universidad Nacional Autónoma
de México

M.^a Dolores García Díaz
Traductora Especializada en Ciencias

© MCMLXXXVI Edición Original John Wiley & Sons, Inc.

© MCMXC Edición Española Ediciones Centrum Técnicas y Científicas

Paseo de Gracia, 26 — 08007 Barcelona — España

Tel. (93) 301 01 82 — Télex 51 735 exit e — Fax (93) 317 97 01

Reservados todos los derechos. Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

ISBN Edición española obra completa: 84-7841-016-3

ISBN Edición española volumen 6: 84-7841-022-8

ISBN Edición original: 0-471-09754-3

Impreso en España — Printed in Spain

Depósito legal B: 2007-91 (En)

Imprime: HUROPÉ, S.A.

Recaredo, 2-4 Barcelona

Índice de capítulos

NOVENA PARTE (CONTINUACIÓN) GRABACIÓN Y REPRODUCCIÓN DE AUDIO Y VÍDEO

45. Grabación y reproducción de audio, 1365 Katsuaki Tsurushima, Electronic Defense Laboratories y Albert B. Grundy

Sistema Sony de disco compacto, 1366
Grabación en cinta magnética, 1376
Altavoces, 1400

46. Grabación y reproducción de vídeo, 1409 RCA Corporation

Fundamentos de la grabación en cinta
de vídeo, 1410
Reproductor de discos de vídeo, 1421

DÉCIMA PARTE COMUNICACIONES

47. Conceptos sobre comunicaciones, 1435 Yeng S. Kuo, Kwei Tu, Rodger E. Ziemer y William H. Tranter

Conceptos sobre comunicaciones
modernas, 1436
Teoría del ruido, 1440
Teoría de la información, 1451

48. Técnicas de comunicación, 1461

Jack W. Seyl, George W. Raffoul,
Matthew J. Quinn, Jr., K. K. Chow,
Bernard Sklar, Peter Monsen, E. T.
Dickerson

Técnicas de modulación, 1463
Radiocomunicaciones, 1510
Comunicaciones alámbricas, 1539
Sistemas de comunicación por
fibras ópticas, 1552
Técnicas avanzadas para las
comunicaciones digitales,
1571
Comunicaciones por canales
con desvanecimiento, 1601
Análisis comparativo de las
técnicas de comunicaciones,
1610

CAPÍTULO 45

Grabación y reproducción de audio

Katsuaki Tsurushima

Sony Corporation of America
Park Ridge, New Jersey

Electronic Defense Laboratories

Sylvania Electric Products, Inc.
Mountain View, California

Albert B. Grundy

Audio Research Institute
Nueva York

45.1 Sistema Sony de disco compacto

- 45.1.1 Introducción
- 45.1.2 Principios de operación
- 45.1.3 Ventajas del sistema de reproducción CD
- 45.1.4 Futuro del sistema CD

45.2 Grabación en cinta magnética

- 45.2.1 Grabación magnética
- 45.2.2 Cabezas de grabación en cinta
- 45.2.3 Mecanismos de transporte de la cinta

45.2.4 Circuitos electrónicos

45.2.5 Cinta magnética

45.3 Altavoces

- 45.3.1 Introducción
- 45.3.2 Tipos de transductores
- 45.3.3 Tipos de muebles acústicos
- 45.3.4 Filtros separadores de frecuencias
- 45.3.5 Patrones polares
- 45.3.6 Técnicas de medición

45.1 SISTEMA SONY DE DISCO COMPACTO

Katsuaki Tsurushima

45.1.1 Introducción

CD es la abreviatura de *compact disc digital audio system* (sistema digital de audio de disco compacto), anunciado en junio de 1980, como resultado de una investigación conjunta entre Sony y Philips. En este programa, Philips investigó los principios básicos de operación y diseñó el *hardware*. La contribución de Sony fue esencialmente el desarrollo del *software*, incluyendo el método de procesamiento de señales.

Entre las metas de diseño del sistema CD no se previó la inclusión de la intercambiabilidad con sistemas de disco de vídeo. En cambio, el sistema ofrece una serie de ventajas que compensan con creces esa falta de intercambiabilidad. Entre tales ventajas sobresalen las del pequeño diámetro del disco (12 cm), larga duración, más de 60 min por lado, fácil manejo, menores costos de producción y la posibilidad de hacer también bastante compacto el sistema modular reproductor para utilizar el disco.

Estos méritos se han logrado adoptando la grabación digital directa mediante modulación por codificación de pulsos (PCM, de *pulse-code-modulation*), así como sistemas completamente nuevos de modulación y corrección de errores.

45.1.2 Principios de operación

En el sistema CD se usa un sistema de lectura de señales sin contacto con un láser de semiconductor. El diseño sin contacto ofrece una ventaja definitiva sobre los sistemas con contacto, como, por ejemplo, los de capacitancia variable y los mecánicos. Esto es, ni la aguja ni el disco se desgastan, no importa cuántas veces se utilice el disco, lo cual asegura una alta calidad y confiabilidad a largo plazo.

Sin embargo, no fue nada fácil incorporar el transductor láser a una unidad compacta. También se encontraron diversas dificultades técnicas en el diseño de los servosistemas rotacionales de velocidad lineal constante (CLV, de *constant linear velocity*), de enfoque y de seguimiento de las pistas, ya que estos sistemas requieren complicados dispositivos ópticos en sus servocircuitos.

A fin de superar estos problemas y hacer del sistema CD una realidad práctica se requirieron tecnología innovadora e ingeniería creativa.

Estructura y especificaciones del disco

Las especificaciones más importantes y la estructura del disco se muestran en la tabla 45-1 y en la figura 45-1, respectivamente. Como puede observarse en esta última (b), el haz láser se aplica desde abajo, pasa por una capa transparente de 1.2 mm de grosor, y se enfoca en la superficie que contiene la señal. Cuando se observa desde el haz láser incidente, la superficie de señal se encuentra cerca del lado superior del disco, que está recubierto con una película plástica protec-

tora, extremadamente delgada (10 a 30 μm). Por consiguiente, la posibilidad de que la superficie de señal se dañe física o químicamente es muy remota.

Como se muestra en la figura 45-2, la función de la capa delgada de 1.2 mm es muy importante. Su índice de refracción (n) es de 1.5. El punto producido por el haz láser sobre la superficie del disco tiene 0.8 mm de diámetro, pero refractado por la capa transparente se reduce hasta 1.7 μm en la superficie de señal. Esto significa que un poco de polvo o una rayadura sobre la superficie del disco mide en realidad sólo la millonésima parte en la superficie de señal. De hecho, cualquier partícula de polvo o rayadura menores de 0.5 μm son insignificantes y no inducen a error en la lectura de la señal. Así, el disco compacto no requiere una manipulación tan delicada como los discos analógicos ordinarios.

En la superficie de señal bajo la capa transparente, el disco contiene una serie de picaduras minúsculas* impresas hacia fuera de la circunferencia interna, con separación de 1.6 μm . Esta separación significa que existen más de 20 000 pistas compuestas de tales picaduras en un área de señal de 33 mm. Dicho de otra forma, es posible empacar 60 pistas en la separación entre pistas de 100 μm de un disco analógico ordinario y aproximadamente 30 pistas en un trozo de cabello delgado.

La precisión de la separación de las pistas del disco compacto en realidad es igual a la del dibujo de los patrones de microcircuitos integrados a gran escala.

La singularmente estrecha separación de las pistas del disco compacto requiere, por supuesto, una extraordinaria precisión del orificio central y de la excentricidad de las pistas. Por este motivo fue necesario desarrollar técnicas de corte y estampado de alta exactitud. Además, tuvo que concebirse un avanzado servosistema de seguimiento de pistas para asegurar un seguimiento exacto compensando la ligera desviación (del orden de 0.1 mm) en el tamaño del orificio central.

Sistema óptico de lectura de señales

En datos digitales, las picaduras y la superficie reflectora del disco compacto corresponden a unos y ceros. En otras palabras, diversa información digital está representada por la presencia y longitud de las picaduras sobre la superficie de señal. Toda la superficie de señal se trata con un recubrimiento reflector de aluminio.

La altura de cada picadura es microscópica: 0.11 μm . Esta cifra se encuentra muy próxima, aunque es ligeramente inferior, al cociente que se obtiene al dividir la longitud de onda del láser ($\lambda = 780 \text{ nm}$) entre 4 y después entre el índice de refracción ($n = 1.5$). Una de las razones para elegir esta altura fue facilitar el uso de la señal de detección de errores de seguimiento por medio de un método simétrico o de contrafase (en *pastipull*). La segunda razón puede explicarse de la manera siguiente.

Supóngase que la altura de la picadura es de $\lambda/4$. Entonces se crea una diferencia de fase igual a $\lambda/2$ o

*Técnicamente se denominan protuberancias o abolladuras, pero aquí se utilizará el término más general "picaduras".

Tabla 45-1. Especificaciones del reproductor de disco compacto Sony

Variable	Especificación
Disco	
Tiempo de ejecución	Aprox. 60 min. sobre cada lado
Rotación	En sentido antirrotatorio cuando se observa desde la superficie de lectura
Velocidad de rotación	1.2-1.4 m/s
Separación entre pistas	1.6 μm
Diámetro	120 mm
Grosor	1.2 mm
Diámetro del orificio central	15 mm
Área de grabación	46-117 mm
Área de señal	50-116 mm
Material	Cualquier material transparente con índice de refracción de 1.5, tal como el plástico acrílico
Longitud mínima de picadura	0.833-0.972 μm (1.2-1.4 m/s)
Longitud máxima de picadura	3.05-3.56 μm (1.2-1.4 m/s)
Profundidad de picadura	Aprox. 0.11 μm
Ancho de picadura	Aprox. 0.5 μm
Sistema óptico	
Longitud de onda estándar	$\lambda = 780 \text{ nm}$ (7 800 Å)
Profundidad focal ($\lambda/\text{NA} \leq 1.75 \mu\text{m}$; NA, apertura numérica)	$\pm 2 \mu\text{m}$
Formato de la señal	
Número de canales	Dos (la grabación en cuatro canales también es posible a dos veces la velocidad actual de rotación)
Cuantización	Cuantización lineal de 16 bits
Sincronización de cuantización	Concurrente para todos los canales
Frecuencia de muestreo	44.1 kHz
Velocidad de bits por canal	4.3218 Mb/s
Velocidad de bits de datos	2.0338 Mb/s
Razón de bits de datos/canal	8:17
Código de corrección de error	CIRC (con 25% de redundancia)
Sistema de modulación	EFM

180° entre el haz reflejado de una picadura y el reflejado de una superficie adyacente, provocando que estos dos haces interfieran y se cancelen entre sí.

En otras palabras, dado que el punto del láser es igual o mayor que cualquier dimensión de picadura, el haz reflejado se modula de manera proporcional al cociente de la picadura sobre el punto, que puede ser detectado por el fotodiodo y convertirse en una señal eléctrica (fig. 45-3).

Captación óptica

En la figura 45-4 se presenta un ejemplo de la forma en que se estructura un transductor óptico. A fin de lograr un enfoque preciso de 1.7 μm de diámetro, fue necesario utilizar como fuente lumínica un láser (dispositivo que produce luz coherente, compuesta de una sola frecuencia [monocromática] y de alta intensidad). El diodo láser que emite el haz láser se coloca en el punto focal de una lente colimadora con distancia focal relativamente grande.

El haz láser que pasa por la lente colimadora se convierte en un haz paralelo, que se enfoca sobre la superficie de señal haciéndolo pasar por la lente objetivo subsecuente, que en este caso es convexa y se caracteriza por tener aproximadamente el mismo diámetro y distancia focal de 4 mm. Sin embargo, a pesar de ser tan pequeña, la lente es tan precisa como las utilizadas en los microscopios más avanzados. Además posee una brillantez sobresaliente y ausencia de aberraciones.

Entre la lente colimadora y la lente objetivo se encuentra el divisor de polarización del haz, que es una especie de prisma que contiene membranas dieléctricas y sirve para dirigir el haz del diodo láser hacia la superficie de señal y el haz reflejado hacia el fotodiodo.

Supóngase que el haz proveniente del diodo láser se polariza horizontalmente. El haz polarizado de esta forma puede pasar a través del divisor de polarización hacia la siguiente placa de un cuarto de onda. Entonces, el haz penetra en la lente objetivo, se enfoca sobre la superficie de señal y es reflejado por el recubrimien-

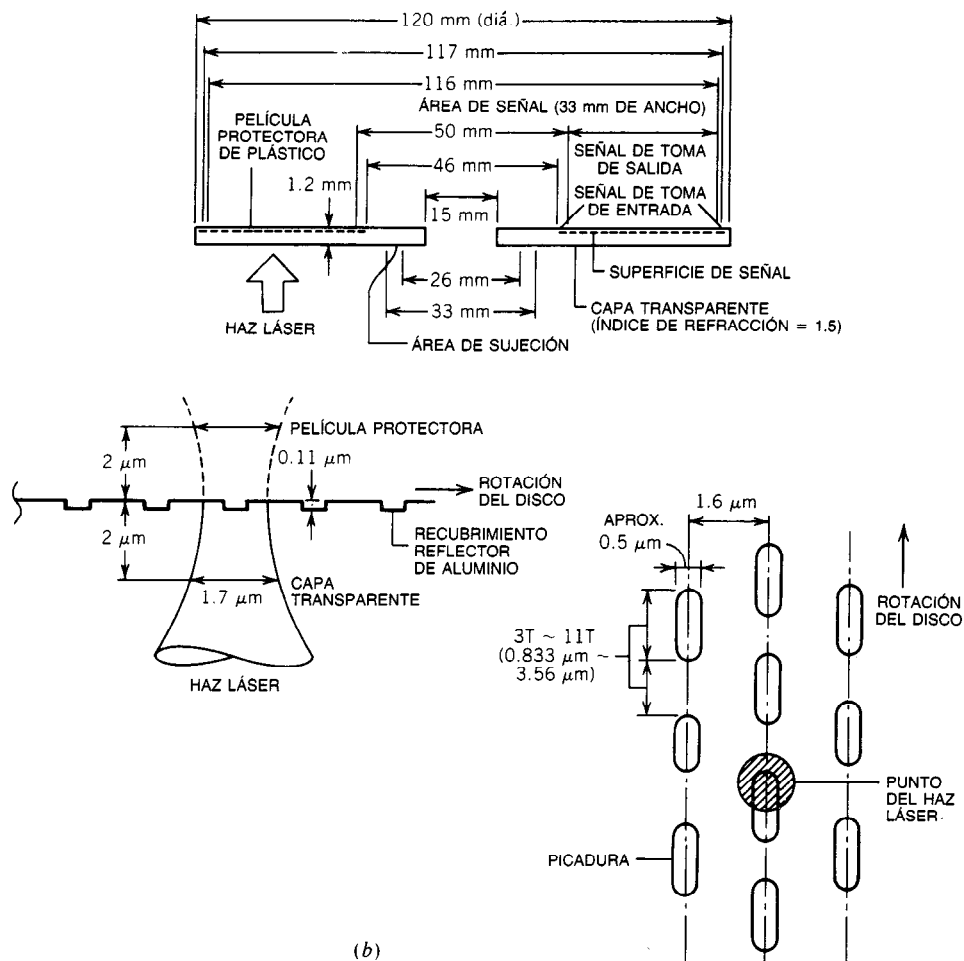


Fig. 45-1. Reproductor del disco compacto Sony: a) unidad; b) diagrama de la estructura del disco.

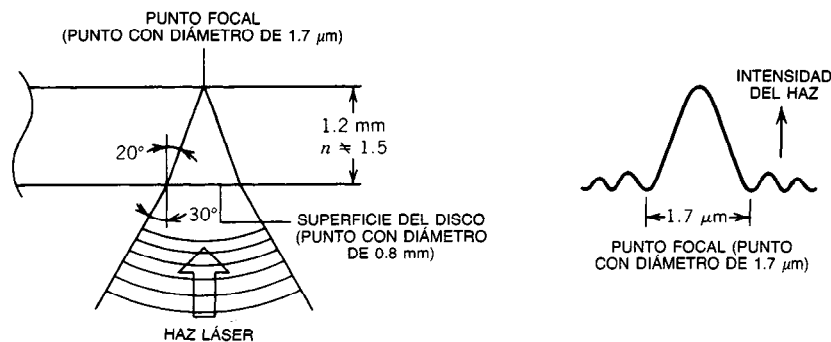


Fig. 45-2. Características del haz láser del sistema digital de audio Sony de disco compacto.

to que se encuentra allí. El haz reflejado regresa a la placa de un cuarto de onda, y al pasar por esta placa por segunda vez se polariza verticalmente.

El divisor de polarización se diseña a fin de no dejar el haz verticalmente polarizado, sino reflejado en dirección del fotodiodo. Así es como puede separarse el haz reflejado del haz incidente para la detección de la señal.

Servomecanismo de enfoque

La profundidad de enfoque de un transductor óptico bien definido, como el empleado en el sistema CD, es de sólo $\pm 2 \mu\text{m}$. Si la superficie de señal se desvía de este intervalo, naturalmente se hace imposible que el transductor detecte las señales. Sin embargo, por desgracia la irregularidad vertical del movimiento del disco a medida que gira es más de 100 veces la profundidad de enfoque, lo que hace necesario un servomecanismo de enfoque que mueva la lente objetivo hacia arriba y abajo, a fin de impedir que la superficie de señal se encuentre fuera de foco.

La figura 45-5a es una ilustración simplificada del paso del haz reflejado. La componente vertical del haz se encuentra concentrada en el punto focal de la lente convexa, y la lente cilíndrica no afecta este haz. Por otra parte, la componente horizontal del haz es refrac-

tada a medida que pasa a través de la lente cilíndrica y crea un punto focal a una distancia más corta. Si el fotodiodo se coloca en el punto de intersección de la componente vertical del haz con la componente horizontal difusa del haz, es posible obtener ahí un patrón circular. Cuando la distancia entre el disco y la lente convexa se incrementa, ambos puntos focales se acercan a la lente, como se muestra por medio de las líneas discontinuas de la figura 45-5a. En este caso el patrón sobre el fotodiodo se hace elíptico, como se muestra en la figura 45-5b. Cuando la relación de posición entre el disco y la lente es al contrario del caso previamente mencionado, es decir, cuando se acorta la distancia entre ellos, el patrón sobre el fotodiodo también se hace elíptico, como se muestra en la figura 45-5b. Por consiguiente, a fin de obtener la señal de control del error de enfoque, primero debe dividirse el fotodiodo en cuatro áreas (A, B, C y D, cuya suma total representa la señal RF), como se ilustra en la figura 45-5c, y restar $B + D$ de $A + C$.

A continuación, es necesario proporcionar la salida del amplificador diferencial al mecanismo de accionamiento de la lente, de forma que la señal de control del error de enfoque $(A + C) - (B + D)$ siempre sea cero. De este modo es posible mantener constantemente enfocada la superficie de señal.

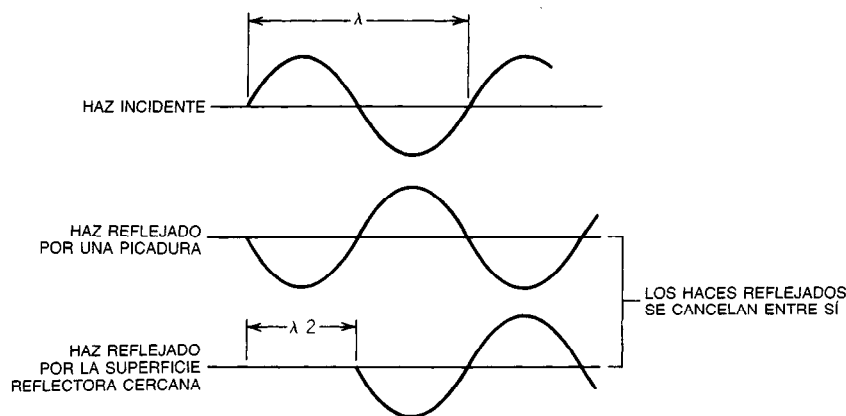


Fig. 45-3. Cancelación del haz.

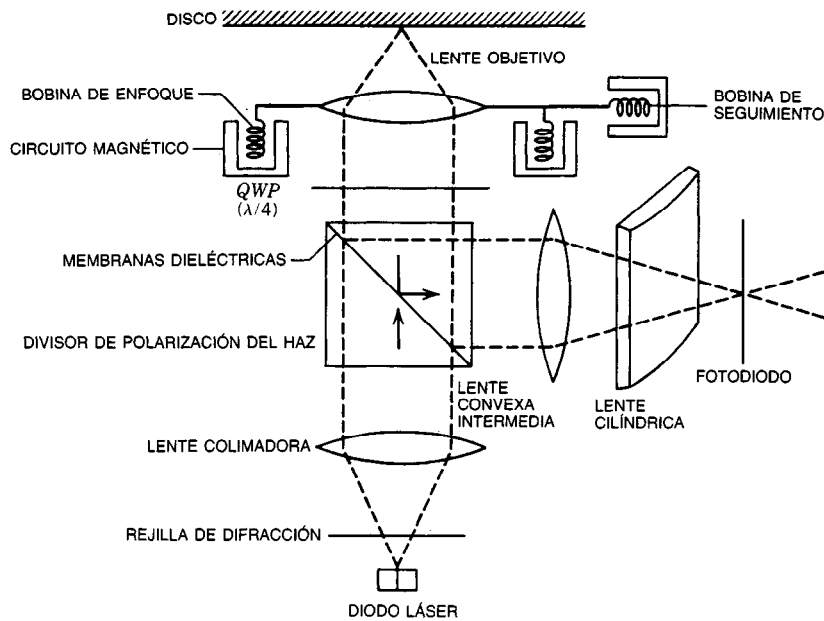


Fig. 45-4. Captación óptica del sistema digital de audio Sony de disco compacto.

Servomecanismo de seguimiento de pistas

Primero es necesario proporcionar patrones extra *E* y *F* a ambos lados del patrón de combinación *A*, *B*, *C* y *D* en el fotodiodo, y debe producirse un punto sobre cada uno de estos tres patrones, como se ilustra en la figura 45-5b. Para este fin se diseña la rejilla de difracción que se muestra en la figura 45-4. Mediante el empleo de su efecto de difracción es posible obtener tres puntos de un solo diodo láser. En la figura 45-6 se muestra la posición relativa de estos puntos sobre la superficie de señal. Si el punto principal *M* se desvía de la pista de señal, se presenta un desequilibrio entre la porción del punto *E* que se superpone a una picadura y la del punto *F*, lo que da por resultado una diferencia en las cantidades de haces reflejados desde dichos puntos. En resumen, el valor que se obtiene al restar *F* de *E* representa la señal de error de seguimiento de pistas, y cuando este valor se mantiene en cero se asegura un seguimiento de pistas correcto. Por consiguiente, con base en el hecho de que el punto focal siempre se presenta en el eje central de la lente, en el sistema CD se emplea un mecanismo de accionamiento exacto que mueve la lente objetivo bilateralmente con respecto a la dirección del haz.

El sistema de control que se acaba de explicar suele conocerse como método de tres puntos. El mecanismo de accionamiento de la lente utilizado aquí es un dispositivo electromagnético parecido a la bobina móvil de un sistema de altavoz. Como un todo, el sistema de accionamiento también se denomina dispositivo de dos ejes, ya que es capaz de ajustar la posición de la lente en dos direcciones, arriba-abajo y derecha-izquierda.

Servomecanismo CLV (de velocidad lineal constante)

CLV significa velocidad lineal constante (*constant li-*

near velocity) y se refiere a un estado en el cual se mantiene una velocidad relativa uniforme entre el disco y el transductor óptico.

Los discos LP ordinarios giran a una velocidad angular constante de 33 1/3 rpm. Comparada con la rotación de los discos LP, la de los discos compactos varía dependiendo de la posición del transductor, a fin de asegurar una velocidad lineal constante de 1.25 m/s. El plato del tornamesa del sistema CD gira a 500 rpm cuando el captador (transductor) sigue la pista cercana a la circunferencia interna. A medida que el captador se desplaza hacia fuera, la velocidad de rotación decrece gradualmente a un mínimo de 200 rpm. El servomecanismo CLV está diseñado para controlar la velocidad ajustando el sincronizador de cuadro impreso en el disco y la frecuencia del oscilador de cristal de cuarzo integrado en el reproductor.

Los tres servosistemas descritos son muy importantes para el sistema CD. Sólo cuando funcionan conforme a lo previsto es posible detectar correctamente las señales digitales.

Sistemas EFM (modulación ocho a catorce)

En la tabla 45-2 se muestra el formato de cuadro. Las señales de audio de los canales L y R (izquierdo y derecho) se muestrean primero a una frecuencia de 44.1 kHz, y después se convierten a representaciones binarias por medio de un proceso de cuantización lineal de 16 bits. Un cuadro se compone de seis de tales señales muestreadas. En consecuencia, cada ciclo de cuadro es igual a 7.35 kHz, o 136 μ s. Con el sistema EFM (de *eight to fourteen modulation*), la señal muestreada se divide primero en dos símbolos, uno que consta de los ocho bits inferiores y el otro de los ocho bits superiores, y a partir de entonces se realiza mo-

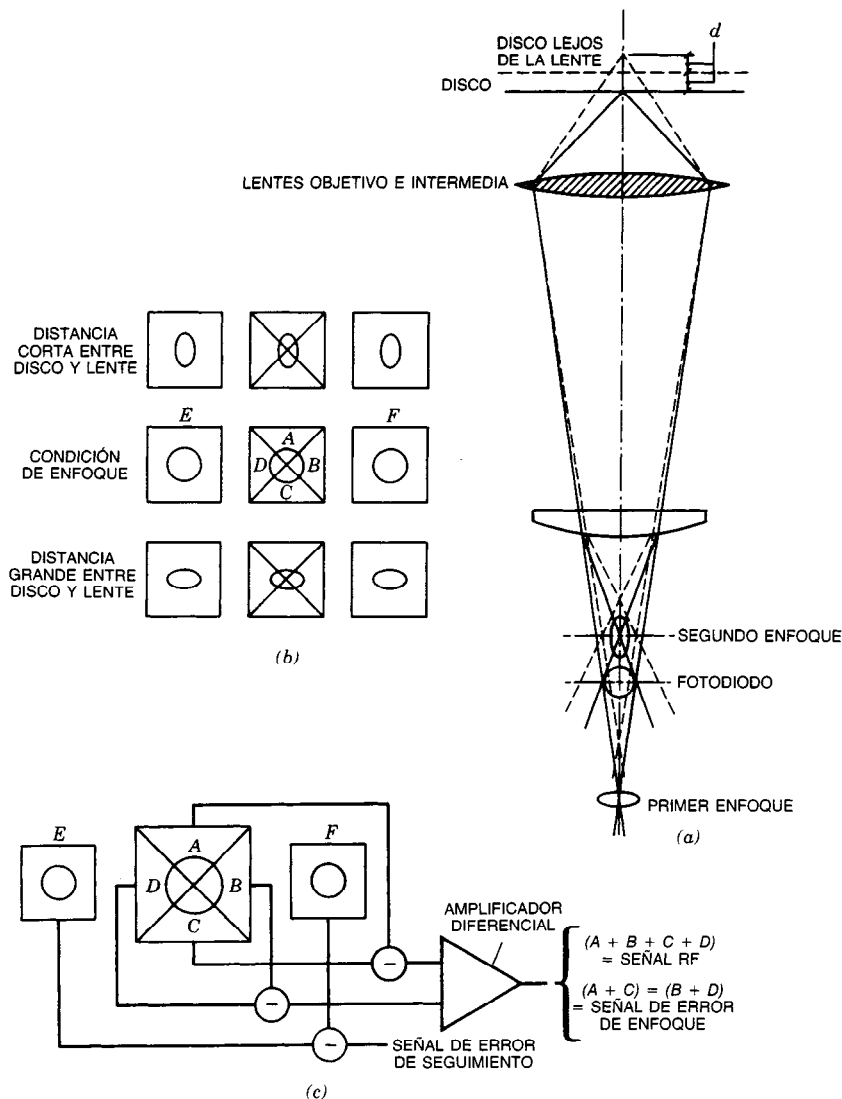


Fig. 45-5. Acción de enfoque del sistema digital de audio Sony de disco compacto: a) paso del haz reflejado; b) patrón en el fotodiodo; c) cálculo de señales.

dulación con base en cada símbolo. La modulación 8 a 14 literalmente implica que cada símbolo de 8 bits se convierte en 14 bits de canal.

Como resultado, la configuración del cuadro pasa a ser como se ilustra en la figura 45-7. Éste es el mejor arreglo entre la más alta densidad de grabación posible y la más fácil extracción posible de bits de reloj.

En notación binaria, 8 bits pueden ofrecer 256 posibles combinaciones de código. Análogamente, 14 bits son capaces de producir $2^{14} = 16\,384$ diferentes patrones de código. Supóngase que de tales patrones se eligen sólo aquellos en los que de manera continua aparecen más de 2 pero menos de 10 ceros. Entonces sólo 267 patrones satisfacen esta condición. En otras palabras, con estos 267 patrones el mínimo ancho de

inversión de una señal NRZ que se invierte en 1 está limitado por la densidad de grabación, y el máximo ancho de inversión está limitado en términos de la extracción de bits de reloj. En la figura 45-8 se indica un ejemplo de código de 8 bits convertido a un código de 14 bits de canal que satisface este requisito. Obsérvese que el intervalo de tiempo original T se amplió a $1.5 T$ en el código convertido. Los dos bits de canal marcados con X en el diagrama son bits de conexión insertados entre patrones de código. En este caso, tales bits corresponden a ceros.

En la aplicación real se agrega un bit más a cada patrón de código. El número binario, 1 o 0, por asignar a este bit extra depende del contenido del código de los patrones antecesor y sucesor. Por decirlo así, se deter-

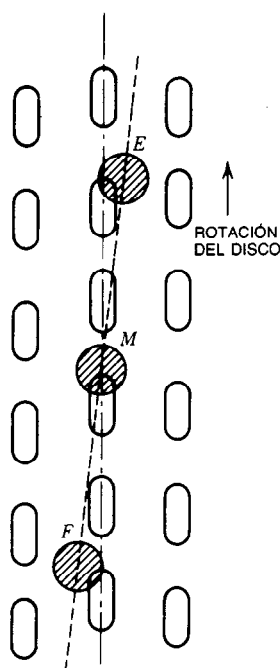


Fig. 45-6. Seguimiento de pistas.

mina de modo que sea posible suprimir la componente de CD que afecta la relación señal-ruido de la señal de error del servomecanismo.

Como ya se ha visto, un símbolo de 8 bits se transforma en 17 bits en el extremo. Sin embargo, en la desmodulación se ignoran los tres bits extra, y los restantes 14 bits de datos se someten a desmodulación.

Éstos son los principios básicos de operación de la EFM. En la práctica, todavía se eliminan 11 de los 267 patrones, por lo que el total de patrones de código memorizados en la memoria sólo de lectura (ROM, de *read only memory*) para su uso es de 256.

Corrección de errores CIRC

Una parte imprescindible del diseño de un sistema de disco digital es un esquema efectivo de corrección

de errores. En términos generales, existen dos tipos de errores de código. El primer tipo son los errores aleatorios, provocados en la producción del disco por imprecisiones en recubrimiento fotorresistivo, corte, etcétera. El segundo tipo son errores inducidos por el uso, resultantes de rayaduras y huellas dactilares en la superficie del disco.

El CIRC (código de intercalación de Reed Solomon, de *cross interleave Reed Solomon code*), creado originalmente por Sony, puede abordar perfectamente ambos tipos de errores. Incluso es capaz de corregir un error externo de más de 4 000 bits. Además, en el sistema CD se utiliza un poderoso método de restauración de código denominado superestrategia. En la figura 45-9 se muestra gráficamente su capacidad de corrección.

En el sistema CD, tres LSI integran el desmodulador EFM y los circuitos CIRC, y éstos, junto con una memoria de acceso aleatorio (RAM), integran una caja negra. En la figura 45-10 se muestra el diagrama de bloques del sistema CD.

45.1.3 Ventajas del sistema de reproducción CD

En la tabla 45-3 se muestra una comparación entre los sistemas de reproducción LP y CD, donde se observa en las especificaciones que este último es muy superior al primero en todos los aspectos de intervalo dinámico, distorsión, respuesta de frecuencia, vibración del tono y gímoteo o efecto de lloro. En especial, el sistema CD presenta un valor notablemente amplio de 90 dB de intervalo dinámico a lo largo de todo el espectro audible de frecuencias. En contraste, el intervalo dinámico del sistema LP es a lo sumo de 70 dB. En los intervalos de baja y alta frecuencia, este valor es aún más bajo, de modo que sólo se informa un intervalo dinámico de 40 a 50 dB. La distorsión armónica del sistema CD nunca es peor del 0.01 %, que es inferior a la centésima parte de la del sistema LP. Además, el gímoteo y la vibración de tono simplemente son demasiado bajos para ser medidos. Esto se debe a que, en la reproducción, los datos digitales se almacenan de una vez en una RAM y después se liberan en forma secuencial perfecta y uniforme, determinada por un preciso reloj de cuarzo de referencia.

Tabla 45-2. Formato de cuadro del sistema digital de audio Sony de disco compacto

Bits	Bits de datos	Bits de canal
Bits de sincronización		24
Bits de control/indicación	$1 \times 8 = 8$	$1 \times 14 = 14$
Bits de datos	$12 \times 2 \times 8 = 192$	$12 \times 2 \times 14 = 336$
Bits de corrección de error	$4 \times 2 \times 8 = 64$	$4 \times 2 \times 14 = 112$
Bits de conexión y bits para supresión de bajas frecuencias		$34 \times 3 = 102$
Total	261	588

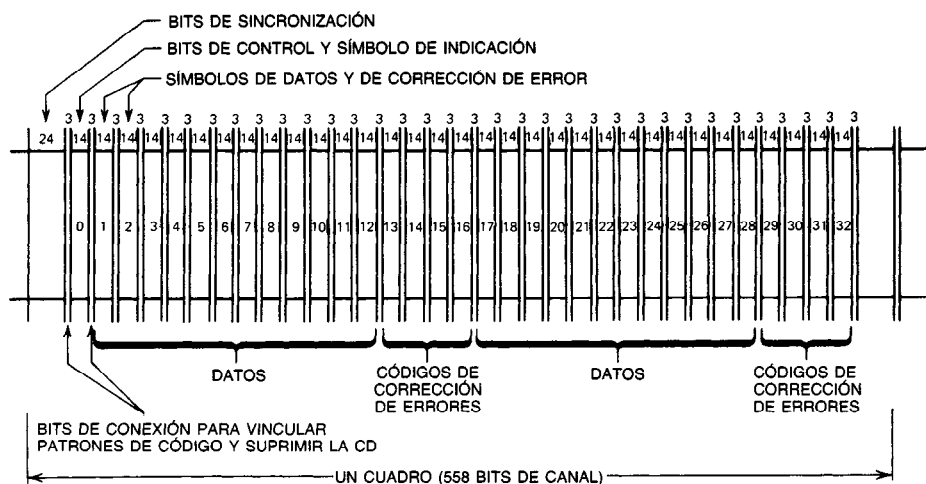


Fig. 45-7. Configuración del cuadro del sistema digital de audio Sony de disco compacto.

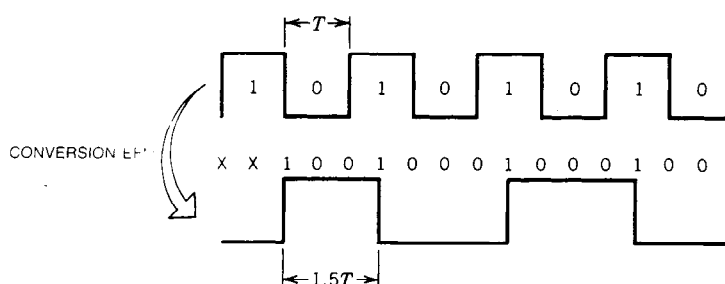


Fig. 45-8. Sincronización.

El reproductor CD tiene otras varias características útiles. Aunque el acceso manual a una pieza musical (corte) del disco es prácticamente imposible debido a que la separación entre pistas es de apenas $1.6 \mu\text{m}$, el reproductor de CD permite total control electrónico a través de la unidad de procesamiento central (CPU, de *central process unit*) integrada y los ocho bits de control insertados después de los bits de sincronización ("sync") en el disco. Los bits de control equivalen aproximadamente al 2.7 % de todos los códigos digitales grabados en el disco. Actualmente, empleando dos de estos bits de control, la información que se describe enseguida puede grabarse para mayor comodidad de operación.

Señales de entrada y de salida. La señal de entrada se encuentra enfrente de donde empieza el área de señal, mientras que la señal de salida aparece después del final de dicha área. Estas dos señales se utilizan para controlar el movimiento del transductor óptico.

Tabla de contenido. En cada área de entrada se encuentra escrita la información de tiempo acerca de los códigos de control, incluyendo el instante de inicio de cada selección, así como el número total y el tiempo de

ejecución de las selecciones, que es posible leer antes de tocar el disco para fines de búsqueda de programas, etcétera. También es posible exhibir de forma óptica tal información.

Códigos de control. Estos códigos diferencian entre grabación de dos y cuatro canales. También son capaces de detectar si se ha efectuado preacentuación (preénfasis) en alguna grabación en particular. Además, incluso pueden cambiar de manera automática las conexiones del circuito del aparato reproductor conforme a las condiciones detectadas del disco.

Señal de inicio de música. La señal de inicio de música se encuentra en el espacio en blanco entre selecciones. Por consiguiente, mediante la detección de esta señal es posible localizar cualquier selección. Incluso cuando el final de una selección se superpone al inicio de otra, es posible insertar la señal durante un mínimo de 2 s frente a la segunda selección, de modo que ambas pueden reconocerse como dos piezas separadas.

Número e índice de la pista. Las selecciones grabadas en el disco pueden numerarse del 1 al 99, y cada barra en una pieza también puede direccionarse hasta el 99.

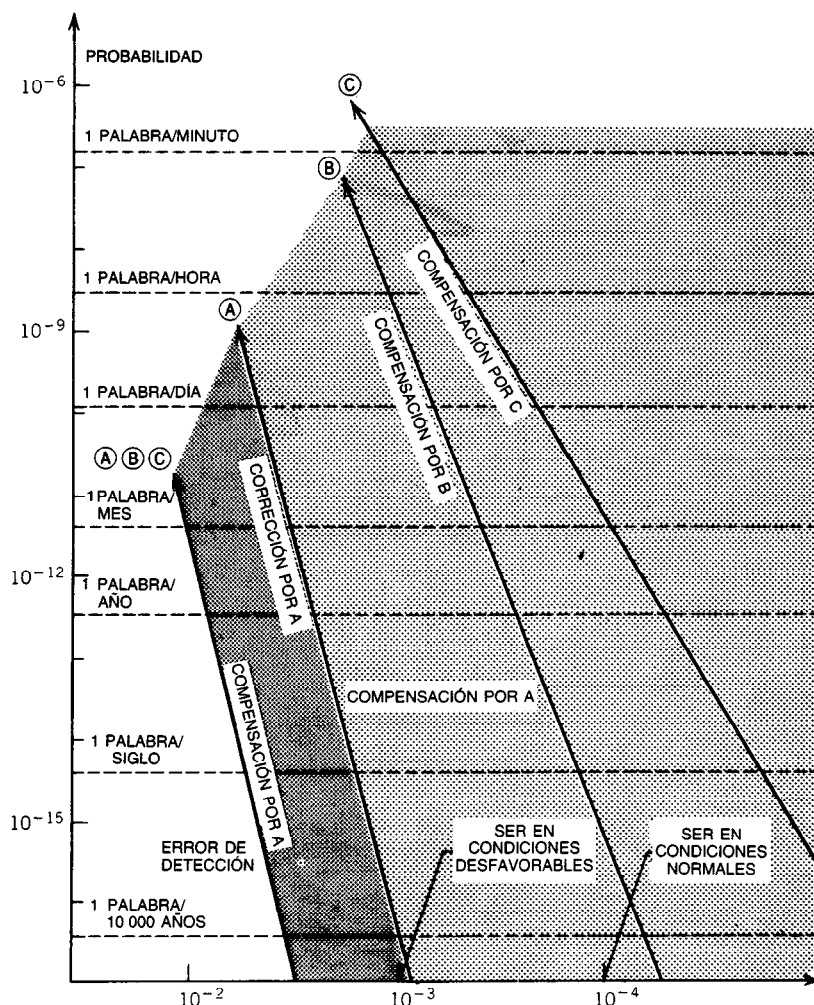


Fig. 45-9. Capacidad de corrección de "superestrategia", método de restauración de código desarrollado por Sony y utilizado en el sistema digital de audio Sony de disco compacto. A = superestrategia; B = sistema estándar de corrección de errores; C = sistema simplificado de corrección de errores. SER = probabilidad de error por símbolo (8 bits) antes de corregir.

Código de tiempo. Con la función código de tiempo es posible establecer el lapso de tiempo a partir del inicio de cada selección en minutos, segundos y 1/75 de segundo. Durante el espacio en blanco entre selecciones, el tiempo se cuenta de manera regresiva.

Como ya se dijo, los bits de control pueden emplearse de diversas formas. No sólo permiten localizar con exactitud y rapidez cualquier porción de música, sino que también permiten la reproducción aleatoria en cualquier orden establecido. Y a pesar del gran tiempo de ejecución de una hora por lado, es posible el acceso al punto deseado de forma casi instantánea.

Este acceso instantáneo es una de las grandes ventajas del sistema CD que no es posible encontrar en ningún otro sistema.

45.1.4 Futuro del sistema CD

El disco compacto y su sistema de reproducción se introdujeron en el mercado simultáneamente hacia fines de 1982. En esa época el precio del reproductor era bastante elevado; sin embargo, en poco tiempo, a medida que madura la industria de partes ópticas, ese precio ha ido disminuyendo a un nivel bastante accesible. En cuanto al disco, se espera que el costo por tiempo de ejecución sea igual al de un disco LP.

Con el actual sistema CD sólo se emplea la cuarta parte de los bits de control, y seis bits permanecen sin utilizar. Esto sugiere que es posible emplear bits de redundancia para grabación, por ejemplo, del título y el texto de la música, para el placer adicional de cantar como acompañamiento de la música CD.

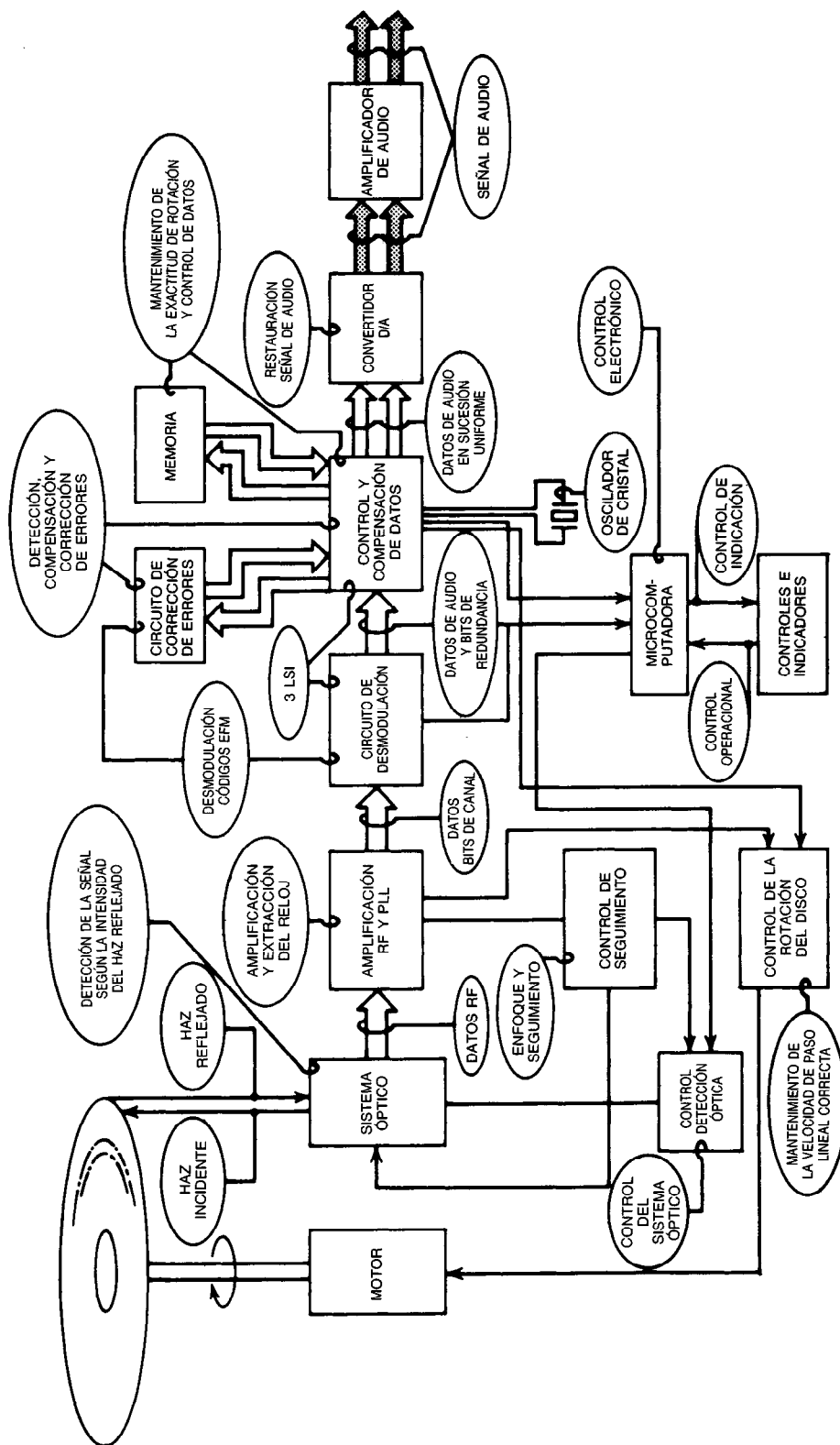


Fig. 45-10. Diagrama de bloques (del disco a la señal de audio) del sistema digital de audio Sony de disco compacto.

Tabla 45-3. Comparación del sistema digital de audio Sony de disco compacto (CD) y el sistema de larga duración (LP)

Variable	Sistema CD	Reproductor LP ordinario
Especificaciones		
Respuesta de frecuencia	20 Hz–20 kHz \pm 0.5 dB	30 Hz–20 kHz \pm 3 dB
Intervalo dinámico	Más de 90 dB	70 dB (a 1 kHz)
Señal-ruido	90 dB (con MSB)	60 dB
Distorsión armónica	Menos del 0.01 %	1-2 %
Separación	Más de 90 dB	25-30 dB
Gimoteo y vibración de tono	Precisión del cuarzo	0.03 %
Dimensiones		
Disco	12 cm (diámetro)	30 cm (diámetro)
Tiempo de ejecución (de un lado)	60 min (máx. 74 min)	20-25 min
Operación-confiabilidad		
Durabilidad del disco	Semipermanente	Respuesta de alta frecuencia degradada después de haberse ejecutado algunas decenas de veces
Durabilidad de la aguja	Más de 5 000 h	500 a 600 h
Operación	Acceso fácil y rápido debido al control por microcomputadora, posible variedad de ejecución programada, mayor resistividad a la vibración externa	Requiere ajuste de la presión de la aguja, es afectado con facilidad por vibraciones externas
Mantenimiento	Polvo, rayaduras y huellas digitales casi insignificantes	El polvo y las rayaduras producen ruido

Además de su utilización como componente de un sistema de audio para el hogar, el sistema CD es adecuado para aplicaciones portátiles y en automóviles, ya que incorpora servosistemas precisos, por lo que es altamente resistente a vibraciones. De hecho, con un poco más de elaboración en su diseño óptico es posible que llegue a comercializarse un "walkman" CD. Además, a diferencia de los reproductores de discos ordinarios, el sistema CD jamás experimenta chillidos, incluso cuando se coloca sobre la caja de un altavoz.

2. Un medio magnetizable que sea conformado por el patrón magnético y lo retenga.
3. Un dispositivo que sea capaz de detectar dicho patrón magnético y convertirlo de nuevo en la señal eléctrica original.

Estos tres elementos adquieren la forma física de la cabeza de grabación, la cinta magnética y la cabeza de reproducción o reproductor. Si se añade algo de amplificación electrónica y un manipulador mecánico de la cinta, se obtiene una grabadora de cinta magnética básica.

45.2 GRABACIÓN EN CINTA MAGNÉTICA*

Cuerpo técnico de los Electronic Defense Laboratories de Sylvania Electric Products, Inc.

45.2.1 Grabación magnética

Para realizar una grabación magnética y después reproducirla se requieren tres elementos básicos:

1. Un dispositivo que sea capaz de responder a una señal eléctrica y crear un patrón magnético en un medio magnetizable.

Grabación y reproducción directas

Una cabeza de grabación es parecida a un transformador con un solo devanado. La corriente de la señal fluye en el devanado, produciendo un flujo magnético en el material del núcleo. Dado que se comporta como cabeza de grabación, el núcleo se elabora en forma de anillo cerrado, pero a diferencia del núcleo del transformador, el anillo cuenta con un pequeño hueco no magnético o entrehierro, que al ser derivado mediante la cinta magnética hace que el flujo se desvíe alrededor de dicho entrehierro a través de la cinta, completando la trayectoria magnética por el material del núcleo. La cinta magnética es sencillamente una tira de plástico sobre la cual se han depositado de manera uniforme pequeñas partículas de material magnético. Cuando la cinta se mueve a través del entrehierro de la cabeza de

*Reimpreso con permiso de la General Telephone and Electronic Corporation, Mountain View, California.

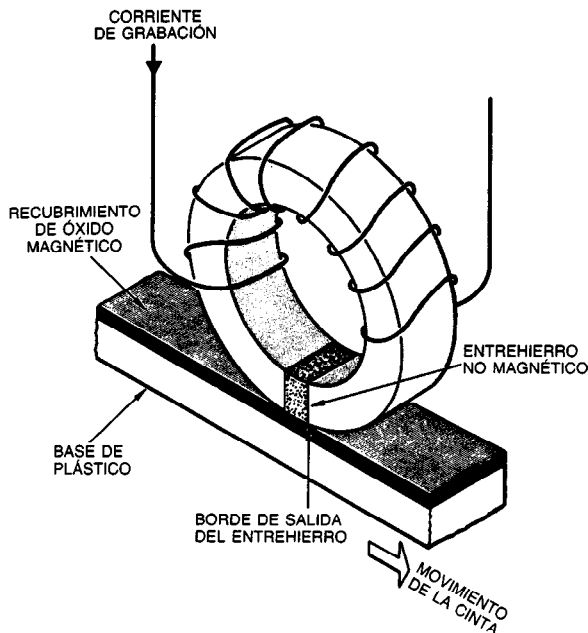


Fig. 45-11. Diagrama simplificado del proceso de grabación magnética.

grabación, el material magnético u óxido se somete a un patrón de flujo que es proporcional a la corriente de la señal en el devanado de la cabeza. A medida que se aleja del entrehierro, cada pequeña partícula retiene el estado de magnetización que le fue impuesto por el flujo saliente. Así, la verdadera grabación se realiza en el borde de salida del hueco de la cabeza de grabación. En la figura 45-11 se muestra un diagrama simplificado del proceso de grabación.

Para reproducir la señal, el patrón magnético de la cinta se mueve a través de una cabeza de reproducción. De nuevo, un pequeño entrehierro no magnético en el núcleo de la cabeza se deriva por medio del recubrimiento de óxido o magnético de la cinta. Las líneas del flujo magnético se derivan a través del núcleo, y son proporcionales al gradiente magnético del patrón abarcado por el entrehierro sobre la cinta. En este momento, el análisis de la función de reproducción se divide en dos posibles alternativas: el empleo de amplificadores de corriente o de voltaje en los circuitos electrónicos de reproducción. Aquí se considera por separado cada método, pero para ambos se emplea el mismo ejemplo a fin de facilitar la comparación.

Supóngase que la señal por grabar en la cinta es un voltaje de onda sinusoidal descrita por $A \sin(\omega t)$. Tanto la corriente en el devanado de la cabeza de grabación como el flujo ϕ a través del núcleo de esa cabeza serán proporcionales a este voltaje.

Amplificador de voltaje. El voltaje inducido en el devanado de la cabeza obedece a la ley de inducción electromagnética: $e_s = N d\phi/dt$. Es importante observar que el voltaje reproducido no es proporcional a la

magnitud del flujo, sino a su tasa de cambio (fig. 45-12). Si la cinta retiene este patrón de flujo y lo regenera en el núcleo de la cabeza de reproducción, entonces el voltaje en el devanado de dicha cabeza estará dado por:

$$e_{\text{repro}} \propto \frac{d\phi}{dt}$$

donde:

$$\begin{aligned} \frac{d\phi}{dt} &= \frac{d}{dt} A \sin(\omega t) \\ &= \omega A \cos(\omega t) \end{aligned}$$

Por tanto, la cabeza de reproducción actúa como diferenciador y la señal reproducida es en realidad la derivada de la señal grabada y no la señal en sí. Este hecho impone dos limitaciones bastante conocidas al proceso de reproducción directa. La salida de la cabeza de reproducción es proporcional a la frecuencia de la señal, y para mantener la fidelidad de la amplitud es necesario compensar en el amplificador de reproducción un incremento de 6 dB por octava en la salida de la cabeza mediante un proceso denominado ecualización.

Amplificador de corriente. La corriente inducida en el devanado de la cabeza obedece a la ecuación $i_s = n\phi/L$ (fig. 45-13). La corriente correspondiente en el devanado de la cabeza de reproducción es:

$$i_{\text{repro}} \propto \frac{\phi}{L}$$

donde:

$$\phi = A \sin(\omega t)$$

Ya que esto elimina el incremento de 6 dB por octava en la salida de la cabeza (fig. 45-14), los circuitos necesarios para la ecualización sólo deben compensar las pérdidas normales de alta frecuencia que se encuentran en la grabación de la cinta. La atenuación resultante de la señal es mucho mayor para las longitudes de onda corta (altas frecuencias), pero es plana para las frecuencias más bajas. Es posible utilizar un circuito L-C reforzador después del preamplificador en

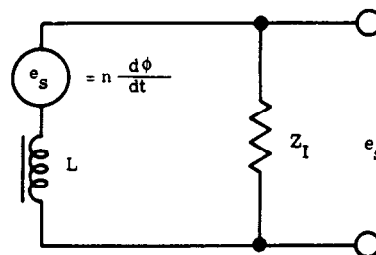


Fig. 45-12. Modelo equivalente de voltaje de la cabeza.

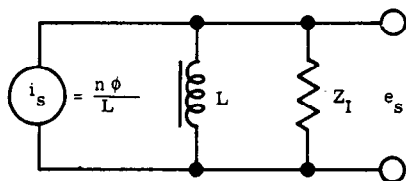


Fig. 45-13. Modelo equivalente de corriente de la cabeza.

las regiones de alta frecuencia a fin de duplicar en efecto el voltaje del circuito de la cabeza y hacer así equivalentes los dos sistemas. Se logra una ligera ganancia en el nivel de la señal en el modelo de corriente, en virtud de que se ha eliminado la atenuación de las frecuencias superiores debida a las pérdidas en la cabeza.

Una comparación global produciría los siguientes datos sobre los amplificadores de corriente:

1. La salida es independiente de la velocidad de la cinta.
2. Se reduce la limitación de ancho de banda de las señales impuesta por la capacidad de entrada.
3. Se minimiza el efecto de las pérdidas en la cabeza sobre la amplitud de la señal.
4. Se elimina el efecto de las variaciones en la inductancia de la cabeza sobre las señales de alta frecuencia.
5. Aunque las características señal-ruido (S/N, de *signal-to-noise*) tanto de los amplificadores de voltaje como de los de corriente son iguales, es posible utilizar conductores con elevada Q en los segundos a fin de obtener mayores relaciones S/N.

Para ambos tipos de amplificación se produce otra limitación a medida que la frecuencia grabada tiende a cero. En algún punto el voltaje de salida de la cabeza reproductora cae por debajo del nivel de ruido inherente al sistema global de grabación, por lo que existe un límite de baja frecuencia en el proceso de grabación directa, por debajo del cual no es posible efectuar la reproducción.

Polarización

Hata el momento, en el análisis se ha supuesto que el medio magnetizable responde linealmente a la fuerza de magnetización de la cabeza de grabación. Como podría esperarse, la perversidad de la naturaleza se hace valer y resulta que tal suposición es errónea. Como otros materiales magnéticos, las partículas depositadas sobre la cinta presentan una característica bastante no lineal cuando son expuestas a una fuerza magnética. En la figura 45-15 se muestra una curva típica de magnetización, o ciclo de histéresis.

H es la fuerza magnetizadora, determinada por el número de vueltas y la corriente en el devanado de la cabeza de grabación. B es la magnetización inducida resultante en la cinta.

Cuando una partícula desmagnetizada en la cinta se aproxima al entrehierro de la cabeza de grabación, no lleva magnetismo residual (punto 0, en el origen,

fig. 45-15). Suponiendo que un ciclo de la señal grabada a lo largo de la cinta es muy grande en comparación con la longitud del entrehierro, la partícula pasará a través de una fuerza de magnetización esencialmente constante creada por la corriente de grabación. Con referencia a la figura 45-15, tal fuerza (H_R) llevará a la partícula por la curva 0A hasta el punto R, en el centro del entrehierro. Cuando la partícula se aleja del entrehierro, H cae a cero, pero la magnetización de la partícula seguirá un ciclo de histéresis menor, RB_R , reteniendo una magnetización residual o remanente de B_R . La característica de transferencia de este proceso se muestra en la figura 45-16, y su no linealidad inherente es obvia. Si no se toman medidas correctivas, se obtiene una alta distorsión en la señal.

Por fortuna existen dos segmentos aceptablemente lineales en la curva de la característica de transferencia, uno a cada lado del origen con su centro aproximadamente a la mitad de camino del punto de saturación (fig. 45-16). Si la grabación puede confinarse a una (o a ambas) de estas secciones rectas, es posible obtener una baja distorsión. Es necesario utilizar algunos métodos para "empujar" la función de grabación hacia la región de transferencia lineal. Los primeros diseñadores de grabadoras recurrían como es natural a una polarización de CD producida simplemente agregando una corriente CD constante a la señal, y obtenían operación en uno u otro de los dos intervalos lineales. Con el intervalo relativamente limitado así disponible, las grabadoras polarizadas con CD tienen un intervalo dinámico muy restringido, pero están bastante mejoradas respecto a la grabación con polarización cero. Para poder utilizar ambas secciones lineales de la curva, es necesario implementar algunos medios de conmutación rápida de una a la otra. Esto es exactamente lo que hace una polarización de CA de alta frecuencia. Existen varias teorías acerca de cómo realiza esta función una polarización de CA, pero ninguna explica en realidad todos los aspectos. Una de las teorías más antiguas y todavía ampliamente aceptada, sin embargo, es que en virtud de que la polarización en sí no se observa en la salida de la cabeza de reproducción, de-

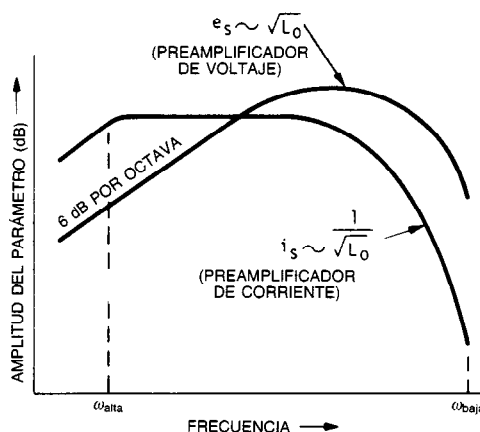


Fig. 45-14. Curvas de intensidad en función del voltaje.

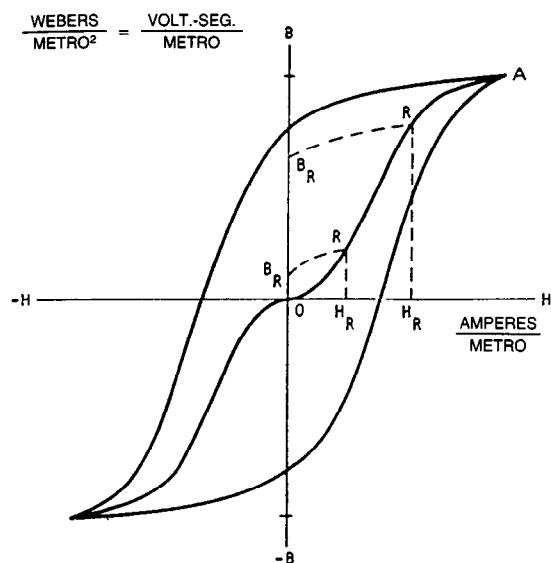


Fig. 45-15. Curva típica de magnetización, o ciclo de histéresis.

bido a su alta frecuencia, su función de conmutación no es detectable y el hueco entre las dos secciones lineales desaparece.

En la figura 45-17 se muestra cómo se graba una señal magnética de baja distorsión. También se logra una reducción adicional en la no linealidad de la misma forma que en la operación de un amplificador en el *push-pull*, ya que las no linealidades se encuentran dispuestas simétricamente alrededor del origen.

Vale la pena mencionar algunas características de la operación de polarización de CA:

1. La polarización y la señal se mezclan (o suman) entre sí de manera lineal. Éste no es un proceso de modulación.
2. La amplitud apropiada para la polarización depende de la característica de transferencia exacta de la cinta y debe ajustarse para abarcar las regiones lineales de centro a centro. Una excesiva polarización reducirá bastante la respuesta de alta frecuencia, mientras que si es insuficiente provocará una mayor distorsión de las frecuencias inferiores.
3. Las frecuencias de polarización no suelen ser críticas, pero deben ser por lo menos 3.5 veces la frecuencia más alta por grabar a fin de minimizar la interacción con los armónicos de la señal.
4. También debe tenerse cuidado de proporcionar una corriente de polarización de onda sinusoidal armónicamente pura, para minimizar los productos de la distorsión.

En la práctica son comunes corrientes de polarización de 1.0 a 20.0 mA, y pueden ser de 5 a 30 veces la corriente de la señal, dependiendo de las características de la cinta y la cabeza.

Respuesta de alta frecuencia

Varios factores se combinan para limitar la respuesta de alta frecuencia de las grabadoras de cinta, pero antes de analizarlos es necesario comprender lo que se entiende por longitud de onda grabada, resolución y densidad de información.

Si se graba una señal sinusoidal, la intensidad magnética de la pista grabada variará sinusoidalmente. La distancia necesaria a lo largo de la cinta para grabar un ciclo completo se denomina longitud de onda trabada, o λ , y es directamente proporcional a la velocidad de la cinta e inversamente proporcional a la frecuencia de

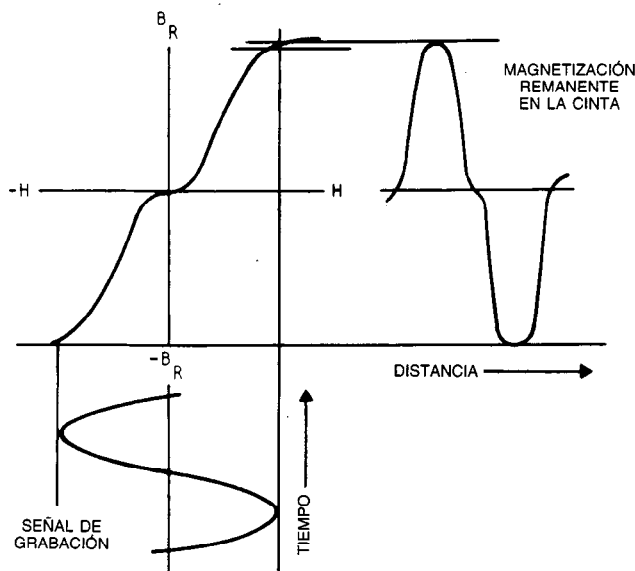


Fig. 45-16. Característica de transferencia cabeza-cinta sin polarización.

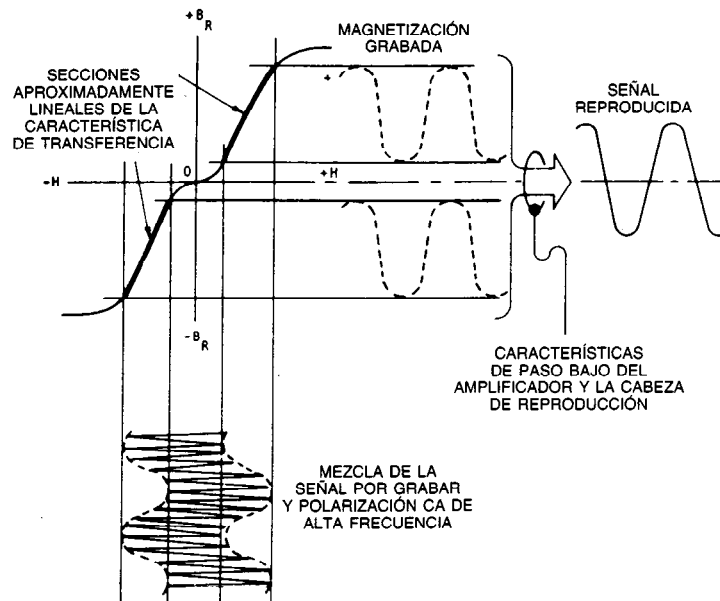


Fig. 45-17. Representación gráfica de la forma en que la polarización de CA transfiere la señal alternadamente desde una sección lineal de la curva a la otra.

la señal. Por ejemplo, en una grabadora específica se observa una respuesta de 60 kHz a 15 pulg/s. Existen otras muchas formas de describir esta respuesta. Al dividir 60 kHz entre 15 pulg se observa que la máquina tiene una *densidad de información* de 4 000 ciclos/pulg. Tal señal tiene una longitud de onda de 0.00025 pulg, que es el límite de *resolución* de la máquina. Es posible utilizar ambas características, la densidad de información y la resolución, para describir la respuesta de una grabadora, independientemente de la velocidad de la cinta, y por tanto definen mejor la capacidad de una grabadora que sólo una especificación de frecuencia a una velocidad dada.

Es posible determinar siete factores que contribuyen a la limitación de alta frecuencia de las grabadoras de cinta:

1. Efecto del entrehierro.
2. Desmagnetización de grabación.
3. Autodesmagnetización.
4. Pérdidas de penetración.
5. Pérdidas en la cabeza.
6. Pérdida por separación.
7. Desalineamiento acimutal.

1. Efecto del entrehierro. Como se muestra en la figura 45-18, la salida de la cabeza de reproducción aumenta con la frecuencia hasta cierto punto, y después disminuye con rapidez hasta cero. La disminución es esencialmente resultado del efecto del entrehierro y ocurre a medida que la longitud de onda grabada (λ) se hace más y más corta hasta que se vuelve igual a la dimensión del entrehierro de reproducción misma. En este

punto no existe un gradiente magnético generado por el entrehierro, y por ello tampoco hay voltaje de salida. Ésta es la restricción independiente más grave sobre la respuesta de alta frecuencia de una grabadora de cinta. En la figura 45-18 se muestra gráficamente el efecto del entrehierro.

- 2. Desmagnetización de grabación.** Ésta provoca un decremento en el nivel de la señal reproducida a las menores longitudes de onda y, como el nombre implica, ocurre durante el proceso de grabación. Normalmente, con las mayores longitudes de onda las partículas en la cinta son accionadas a través de grandes ciclos simétricos de histéresis mediante la polarización de CA mientras se encuentran bajo la influencia del campo de grabación. Estos ciclos son compensados por la señal de grabación, mucho más pequeña; a medida que la partícula abandona el campo se deshace el ciclo, dejando magnetizada la partícula en proporción con la señal. Para longitudes de onda más cortas la señal de grabación puede variar considerablemente a medida que la partícula deja el campo, y resultará una reducción correspondiente en la magnetización remanente.
- 3. Autodesmagnetización.** Esto ocurre en el medio magnético en sí cuando se elimina la fuerza externa de magnetización; es más pronunciada cuando los polos magnéticos se agrupan más y más cerca (longitudes de onda más cortas). En realidad, la autodesmagnetización es probablemente la menos importante de las limitaciones de las altas frecuencias.

4. **Pérdidas de penetración.** Estas pérdidas dependen de la longitud de onda y provocan otra reducción en el nivel de la señal reproducida. La profundidad total (o grosor) del recubrimiento magnético de la cinta se magnetiza a grandes longitudes de onda, pero, a medida que éstas disminuyen, la profundidad de magnetización se reduce y, a longitudes de onda muy cortas, sólo se magnetizan de manera efectiva las partículas de la superficie. Por tanto, las longitudes de onda más cortas afectan pocas partículas, existe menos intensidad en el patrón magnético grabado y la salida reproducida descende.
5. **Pérdidas en la cabeza.** A diferencia de las limitaciones ya analizadas, las pérdidas en la cabeza no dependen de la longitud de onda, pero, como sucede en el caso de cualquier material ferromagnético accionado por CA, están estrechamente relacionadas con la frecuencia. Las pérdidas tanto en el núcleo como en el devanado reducen la corriente efectiva de grabación a altas frecuencias de la señal. Los principales contribuyentes son las pérdidas por histéresis y por corrientes parásitas en el material del núcleo y la capacidad distribuida de los devanados, y por supuesto, aumentan con la frecuencia.
6. **Pérdida por separación.** Este efecto se observa en la señal reproducida como reducciones muy breves y aleatorias (caídas de señal o faltas momentáneas de registro) en el nivel de la señal. En realidad, las caídas no son provocadas por la

grabadora, sino por un contacto imperfecto entre la cabeza y la cinta. La tecnología moderna de cintas ha abatido los problemas de la caída en varios órdenes de magnitud, pero no los ha eliminado. En la grabación de audio, las caídas provocan muy pocos problemas, debido a que el oído humano tiende a integrar las variaciones de corta magnitud, y con este proceso de alisamiento no son discernibles.

7. **Desalineamiento acimutal.** Cuando el entrehierro de la cabeza de reproducción no es exactamente paralelo al de la cabeza de grabación, ocurren pérdidas de longitud de onda corta. Estas pérdidas se definen por medio de:

$$\text{Pérdida (dB)} = 20 \log \frac{\sin[\pi W(\tan A)/\lambda]}{\pi W \tan A/\lambda}$$

donde W = ancho de la pista

A = ángulo de desalineamiento, rad

λ = longitud de onda

Por ejemplo, con un ancho de pista de 37 mpulg y una longitud de onda de 0.25 mpulg ocurre una pérdida de 3 dB para un desalineamiento de 1/6 de grado (0.003 rad). Es obvio que el alineamiento adecuado de la cabeza es importante para reducir las pérdidas por acimut; sin embargo, puede haber otros contribuyentes, como guías de cinta excedidas de tamaño o gastadas, ancho de cinta inferior al normal o alineamiento inadecuado del rodillo de presión

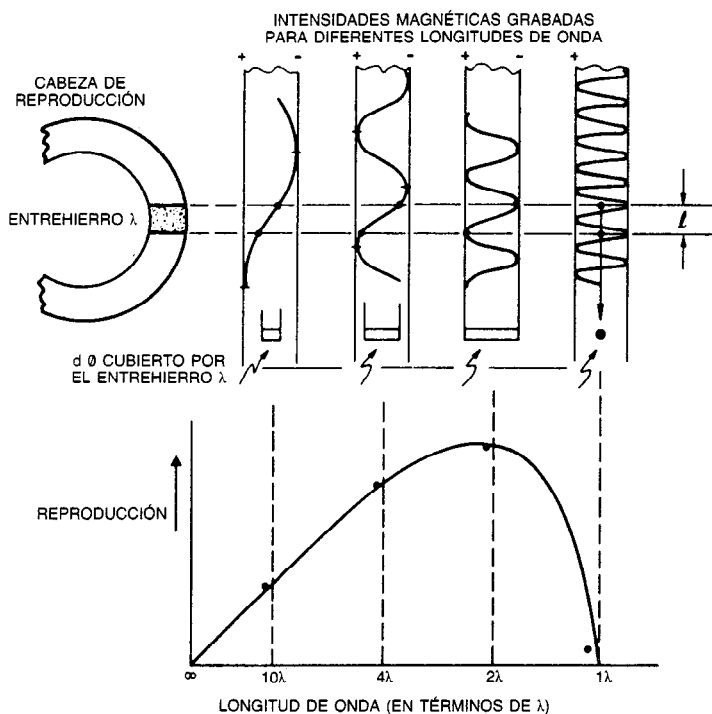


Fig. 45-18. Representación gráfica del efecto del entrehierro.

del transporte. En grabadoras de banda extremadamente amplia, se encuentran integrados a la base del montaje del grupo de cabezas de reproducción unos pernos de ajuste del acimut. Es posible obtener el acimut óptimo de la cabeza de reproducción doblando este ajuste a fin de maximizar la señal reproducida de alta frecuencia. Es posible que no todas las cabezas de un grupo de cabezas reproductoras apiladas sean óptimas para un ajuste de acimut dado.

Otras características

En el análisis de la polarización se demostró que el intervalo lineal de la característica de transferencia se vuelve no lineal gradualmente a medida que la magnetización se acerca a la saturación. Esto da al proceso de grabación lo que se describe atinadamente como limitación "elegante", o en otras palabras, el hecho de que el aumento del nivel de grabación por encima de lo normal incrementará gradualmente la distorsión antes de que ocurra una limitación brusca, provocada por la saturación magnética. Para definir el nivel máximo de la señal que es posible grabar, es necesario entonces establecer la máxima distorsión que es posible tolerar. En la práctica, el máximo nivel especificado de la señal suele estar ligado al 1 % de la especificación de distorsión armónica total (THD, de *total harmonic distortion*).

El intervalo dinámico o relación señal-ruido (S/N, de *signal/noise*) se especifica en decibeles y es la razón entre la máxima señal (para una THD dada) y la mínima señal que es posible grabar; la mínima señal está determinada por el nivel de ruido de todo el sistema en el ancho de banda de interés.

Antes de dejar el proceso de grabación directa, es necesario analizar los problemas asociados con el uso de grabadoras de audio para trabajos de instrumentación. Probablemente todos los lectores estén familiarizados con la gran variedad de grabadoras de audio que existen en el mercado actualmente. La mayor parte de ellas efectúan un excelente trabajo en cuanto a la grabación de la voz y de música, pero no deben confundirse con las grabadoras para instrumentación. Éstas son en realidad un caso muy especial del proceso de grabación en directo. La comparación de las especificaciones de las grabadoras de audio y para instrumentación mostrará con facilidad algunas diferencias. Por ejemplo, las grabadoras de audio registran respuestas hasta de 40 a 50 Hz y valores señal-ruido de 50 a 60 dB, mientras que las grabadoras para instrumentación se detienen a 50 Hz y presentan relaciones S/N de 24 a 38 dB.

Sólo cuando se analizan los objetivos de ambos tipos de grabadora es posible reconciliar estas diferencias. Las máquinas de audio se venden para grabar voz y música y están diseñadas para sacar provecho de las peculiares características de la energía espectral de este tipo de señal. Años de estudio, esencialmente en la industria telefónica, han demostrado que el contenido de energía de tales señales no está distribuido uniformemente sobre el intervalo de frecuencias audibles, sino que se encuentra fuertemente concentrado en el

centro de la banda de audio, con relativamente poca energía en los extremos de alta o baja frecuencia. Como resultado, cada extremo de la banda requiere muy poco intervalo dinámico. El amplificador de grabación del aparato de audio refuerza (preacentúa o preenfatisa) estas señales en el borde de la banda a medida que se colocan en la cinta. Cuando se reproduce la cinta, el amplificador de reproducción de audio opera con ganancia reducida en los extremos de la banda a fin de ecualizar la señal y restaurar fidelidad. Este factor también mejora el nivel global de ruido de la grabadora de audio, ya que es en los extremos de la banda donde el nivel de ruido del sistema es más alto. Con ganancia reducida en dichas regiones se reduce bastante el ruido acumulativo en la banda de transmisión libre. Esta característica de lujo ha hecho ganar a la grabadora de audio menor respuesta de frecuencia y menor nivel de ruido en la señal reproducida, al costo de menor intervalo dinámico (o saturación más pronta) en cada extremo del espectro audible; es decir, a 50 o 14 000 Hz una grabadora de audio puede saturarse con una señal de 12 a 15 dB, aun cuando a 400 Hz (valor al que se mide la relación S/N) el aparato pueda manejar una señal de 55 dB. Este tipo de preénfasis y postecualización ha sido ampliamente aceptado en las grabadoras de audio, tanto para uso en el hogar como profesional, y la National Association of Broadcasters (NAB)* ha adoptado curvas estándar de ecualización de audio. Dichas curvas son parecidas a las de la bien conocida ecualización RIAA utilizada en la industria fonográfica.

Las grabadoras para instrumentación no deben favorecer ninguna parte de su ancho de banda grabada debido a la naturaleza de las señales que deben recibir. Su respuesta de baja frecuencia se cita a 50 cps, de modo que la contribución de ruido agregado por debajo de este punto puede eliminarse. Por supuesto, el ruido agregado por la respuesta de frecuencias más altas se suma al ruido acumulado y reduce el valor S/N. Asimismo, en el esfuerzo por conseguir respuesta de frecuencia más alta se utilizan entrehierros mucho más estrechos en las cabezas de reproducción, con una reducción correspondiente en la potencia de la señal. La definición de máximo nivel de grabación también trabaja en contra del aparato para instrumentación; es decir, en su caso es el nivel que produce el 1 % de THD, mientras que para las grabadoras de audio es el nivel que produce el 3 % de THD.

Como puede observarse, existen muchas diferencias entre las grabadoras de audio y las grabadoras para instrumentación; algunas son evidentes, otras muy sutiles.

Hay muy pocas ocasiones en las que una máquina de audio grabará datos de instrumentación sin peligro. Debe tenerse cuidado en caso de elegir una para el trabajo de grabación.

* La NAB está integrada por miembros interesados en todas las fases de la radiodifusión comercial: AM, FM y TV. NARTB es la denominación anterior de la misma organización; quiere decir National Association of Radio and Television Broadcasters.

Normas de respuesta para el proceso de grabación directa

Como resultado de la necesidad de asegurar la compatibilidad cuando es necesario intercambiar datos entre máquinas en diferentes sitios o de origen diferente, se requiere alguna forma de estandarización.

Se han establecido normas para cuatro grupos de parámetros de grabación directa. Las normas IRIG, establecidas por vez primera a inicios de la década de 1950, son aplicables sólo a los sistemas de cinta de 0.5 y 1 pulg, mientras que los avances tecnológicos han permitido que los fabricantes ofrezcan capacidad de grabación para instrumentación de 4, 7 y 8 canales en cinta de 0.25 pulg. La carencia de normas IRIG para cinta de 0.25 pulg no ha impedido la incorporación de las especificaciones de alto rendimiento aplicables a las nuevas grabadoras. Existen muchos usuarios industriales, médicos, gubernamentales y otros que no requieren las normas de intercambio de IRIG, y que deberían estar informados acerca de la relación precio-rendimiento que estas grabadoras de 0.25 pulg son capaces de proporcionar.

Grabación y reproducción por FM

Después de la Segunda Guerra Mundial, las limitaciones del proceso de grabación directa restringieron gravemente el empleo de las grabadoras de cinta para el trabajo general de instrumentación. Alrededor de 1950, la industria en desarrollo de los aeromisiles y las muchas áreas de prueba del ejército debían grabar una cantidad creciente de datos de ensayos de CD y de muy baja frecuencia. A fin de atender estas áreas, la grabación magnética tenía que proporcionar de alguna manera respuesta de CD, buena linealidad de CD y mejores relaciones señal-ruido.

La industria de la grabación satisfizo esta necesidad con el desarrollo de la grabación por FM. La técnica fue ampliamente aceptada, y la mayor parte de las grabadoras en uso en la década de 1950 fue equipada para operación con FM. Todavía en uso actualmente, el primer esquema de FM utilizó frecuencias centrales y desviaciones adaptadas para grabadoras de 80 a 100 kHz y colocó una sola portadora en cada pista, grabada hasta la saturación sin polarización.

La grabación de datos utilizando una portadora modulada en frecuencia se logra mediante la desviación de la frecuencia de la portadora en respuesta a la amplitud de la señal de datos, y grabándola. Una señal de CD de una polaridad incrementa la frecuencia de la portadora, y la polaridad opuesta la reduce. Una señal de CA incrementa y reduce de manera alternada la portadora por encima y por debajo de su frecuencia central a un ritmo igual a la frecuencia de la señal de datos. En el proceso de reproducción, la inestabilidad de la amplitud de la portadora se elimina esencialmente limitando los cruces por cero, y la señal de datos se reconstruye detectándolos. La señal portadora residual y el ruido fuera de banda se eliminan mediante un filtro pasabajas. Sin embargo, la grabación por FM presenta un problema característico: es extremadamente sensible a las fluctuaciones en la velocidad de la cinta (vibración de tono), debido a que, ya sea en el

modo de grabación o en el de reproducción, dichas fluctuaciones producen modulaciones indeseables (o ruido) de la portadora.

El nivel aumentado de ruido en la señal reproducida y la reducción correspondiente en el intervalo dinámico constituyen el efecto de primer orden de la vibración de tono. Un efecto de segundo orden es la variación de la base de tiempo real del proceso de grabación directa.

En libros de texto es posible encontrar una descripción técnica detallada de la teoría de la FM, pero aquí se incluye un breve análisis de algunos de los factores más relacionados con ella, para ayudar al lector a que comprenda los rudimentos del proceso de grabación por FM. Primero se proporcionan algunas definiciones:

f_c	frecuencia central de la portadora
Δf	máxima desviación de la portadora respecto a f_c
f'	frecuencia de la señal de datos
$f' \text{ máx}$	máxima frecuencia de la señal de datos
$\frac{\Delta f}{f' \text{ máx}}$	relación de desviación, o índice de modulación
$100 \frac{\Delta f}{f_c}$	desviación porcentual

La relación de desviación es uno de los factores más importantes en cualquier proceso de FM. Básicamente, cuanto mayor sea la relación de desviación, más inmunidad contra el ruido tiene el sistema. Sin embargo, en la grabación por FM existen límites prácticos a la relación de desviación, ya que el Δf está restringido por limitaciones del ancho de banda de la grabadora y la $f' \text{ máx}$. debe mantenerse alta a fin de acomodar las señales de datos. Las relaciones de desviación de uso común actualmente varían de 5 en las subportadoras de telemetría por FM a 0.675 en la grabación por FM de banda ancha. En la radiotransmisión en FM, que disfruta de excelente inmunidad contra el ruido, se utiliza una relación de desviación igual a 5 ($\Delta f = 75 \text{ kHz}$, $f' \text{ máx.} = 15 \text{ kHz}$).

La desviación porcentual, $100(\Delta f/f_c)$, es otro factor en la grabación por FM. Cuando se graban sistemas con baja desviación porcentual, tales como las subportadoras de telemetría por FM (7.5 y 15 %), el efecto de la vibración de tono está esencialmente multiplicado con un incremento correspondiente en el ruido. Por ejemplo, si una desviación de frecuencia del 7.5 % corresponde a una señal de entrada del 100 %, una desviación del 1 % provocada por vibración de tono se presentará como $100/7.5 = 13.3 \%$ de señal de ruido. La misma variación de tono impuesta a un sistema con el 40 % de desviación provocará sólo $100/40 = 2.5 \%$ de señal de ruido. Los sistemas con mayores porcentajes de desviación son por tanto menos afectados por variaciones en la velocidad de la cinta, pero ciertas limitaciones en el diseño de circuitos hacen del 70 al 75 % un límite práctico.

terial ferromagnético de alta permeabilidad. Las superficies de las medias secciones, que tienen entrecaras entre sí, se esmerilan y pulen con mucho cuidado, y el material del entrehierro se deposita sobre una de ellas. A continuación se unen los dos núcleos y sus devanados a fin de formar una cabeza.

Se utiliza una cabeza para cada pista de la cinta. En la operación con varias pistas, es necesario ensamblar varias cabezas con protecciones interpistas, a fin de formar un grupo de cabezas apiladas. En un grupo de cabezas es necesaria una precisión extrema para alinear las cabezas, ya que los entrehierros de todas ellas deben caer en una banda de 100 μ pulg. Después del alineamiento, las cabezas se encapsulan. La pila de cabezas se termina por contorneado y pulido a fin de lograr un contacto óptimo cabeza-cinta, y después se monta sobre una base de montaje con labrado de precisión mediante la cual se asegura al mecanismo de transporte de la cinta. Por tanto, el ensamblado completo de la cabeza incluye una base con una superficie de montaje cuyo plano es perpendicular a la línea del entrehierro en la superficie de contacto de la pila de cabezas. Debido a la base de montaje de precisión es posible quitar la pila de cabezas para limpiarla o bien para instalar una nueva sin necesidad de ajuste de fábrica, a la vez que se mantienen mínimas las pérdidas debidas a desalineamiento acimutal (véase el análisis previo).

Cabezas de grabación

Ya se explicó antes que el proceso de grabación se efectúa en el borde de salida del entrehierro de la cabeza de grabación. Debido a esto, la longitud de dicho entrehierro afecta muy poco la respuesta de frecuencia de un aparato. En la mayor parte de las grabadoras se emplea una longitud de entrehierro de grabación de 0.2 a 0.5 mpulg. Sin embargo, la precisión del borde del entrehierro es de suma importancia para la capacidad de alta frecuencia de la grabadora, por lo que se tiene especial cuidado a fin de obtener un borde afilado bien definido.

La combinación ideal cabeza de grabación-amplificador de grabación establecerá un patrón de flujo de igual amplitud en la cinta para un nivel de señal dado a través de todo el intervalo de frecuencias de la grabadora. Ésta sería ciertamente una grabación de "flujo constante". En realidad, con las grabadoras modernas se obtiene un rendimiento bastante aproximado a esta condición ideal por medio del empleo del preénfasis de grabación. Las pérdidas en la cabeza y en alguna medida las pérdidas de desmagnetización de grabación se compensan con el empleo de una característica "reforzadora" de alta frecuencia en el amplificador de grabación. Sin embargo, no se emplea preénfasis adicional, ya que la condición de flujo constante es óptima para el mejor intervalo dinámico y mínimos productos de distorsión.

Cabezas de reproducción

La respuesta de alta frecuencia de una grabadora de cinta se determina por la longitud del entrehierro y la calidad de sus cabezas de reproducción. Para que un

aparato reproduzca una señal de 2 MHz a 60 pulg/s (ips), debe "ver" longitudes de onda grabadas sobre la cinta tan cortas como 30 μ pulg, y esto requiere que la longitud del entrehierro de la cabeza de reproducción sea de aproximadamente 10 μ pulg. Además, el entrehierro debe estar claramente definido y poseer una profundidad adecuada a fin de resistir el efecto abrasivo de la cinta. Las características de desgaste del material del entrehierro y el del núcleo deben ser parecidas a fin de que el entrehierro preserve su definición. La producción en serie de varias de tales cabezas en una sola pila de cabezas es uno de los mayores logros industriales actuales en Estados Unidos. El precio de dicha pila de cabezas de alta resolución puede ser de varios miles de dólares.

El esfuerzo por obtener longitudes de entrehierro más y más cortas tiene una desventaja. Desafortunadamente, a medida que disminuyen dichas longitudes también decrece el voltaje de salida del devanado de la cabeza, reduciendo de esta forma la razón señal-ruido de la grabadora. Uno de los primeros síntomas de desgaste excesivo de la cabeza es el aumento de su salida a frecuencias de señal más bajas. Esto resulta del alargamiento del entrehierro. Sin embargo, aun con tales condiciones adversas, las grabadoras actuales obtienen relaciones señal-ruido de 38 dB e incluso mejores, a la vez que densidades de información de 30 000 a 33 000 ciclos por pulgada.

Cuidado y duración de la cabeza

El cuidado y la duración de las cabezas de las grabadoras de cinta son completamente interdependientes. Algunos de los factores que pueden afectar la duración de la cabeza durante el uso normal son los siguientes:

1. Limpieza de la cinta, el mecanismo de transporte y el ambiente en que opera el equipo.
2. Procedimientos de mantenimiento que implican la comprobación de la cinta, el seguimiento de pistas, etcétera.
3. Capacidad abrasiva de la cinta empleada.
4. Disolventes utilizados para limpiar las cabezas.

La limpieza en y alrededor del área de la cabeza es de la mayor importancia en todos los aparatos para instrumentación. Las partículas de polvo no sólo pueden convertirse en una seria amenaza para la "captura" de datos en términos de pérdidas por separación, sino que también pueden volverse minúsculos raspadores, excavadores y cortadores de las superficies de la cabeza y de la cinta cuando se encuentran entre ellas. A fin de mantener bajas las pérdidas por separación se requieren elevadas presiones de cinta a cabeza, pero tales presiones no sólo incrementan los efectos abrasivos de la cinta, sino que también aplastan cualquier partícula de impureza contra el material de la cabeza o de la cinta.

Las operaciones de mantenimiento de casi todos los aspectos de la parte mecánica de la grabadora afectarán el área de la cabeza de una forma u otra. Las tensiones de la cinta deben mantenerse conforme a las especificaciones a fin de asegurar el mejor arreglo en-

tre desgaste y rendimiento. Es necesario tener cuidado de no tocar las cabezas con cualquier objeto metálico o duro que las pudiera rayar, excavar o magnetizar.

La grabación y reproducción de longitudes de onda cada vez más cortas demandan un estrecho contacto entre las partículas de óxido y el entrehierro de la cabeza a fin de eliminar las pérdidas por separación. El resultado han sido mayores tensiones y superficies de cinta más abrasivas, y ambas tienden a incrementar el desgaste de la cinta. Las partículas gamma del óxido férrico son muy afiladas y duras, y se parecen a partículas extremadamente finas de papel de lija. Los aglomerantes de cinta empleados en cintas de banda ancha son también más lisos y duros, por lo que forman una base firme para cada partícula rayadora. Si no es necesario contar con altas densidades de información, no se requieren cintas de primera calidad.

Para obtener buenas grabaciones es necesario que las cabezas estén limpias. Sin embargo, no es seguro utilizar cualquier disolvente que parezca disolver los residuos dejados por la cinta. El material empleado para retener los núcleos de las cabezas en su ensamblaje algunas veces se ablanda con disolventes, tales como el tolueno, metil-etil-cetona o xileno. Si existen dudas acerca de la construcción de la cabeza, debe emplearse sólo alcohol, nafta, freón TF, gasolina o incluso combustible para aviones a reacción. Probablemente el freón TF es el mejor de todos los limpiadores, y ya existe en el mercado en presentación de bote en aerosol para uso y almacenamiento convenientes. La generalidad de los limpiadores de cabezas disuelven grasas lubricantes y aglomerantes de cinta, por lo que es necesario utilizarlos con cuidado, en especial alrededor de los cojinetes y de la cinta. Un hisopo de algodón es una herramienta de limpieza desechable apropiada para la mayor parte de los requerimientos de limpieza.

Hay que limpiar todas las partes de los mecanismos de transporte que entran en contacto con la cinta. La acumulación de óxido puede provocar una degradación de la respuesta de alta frecuencia, grabación y reproducción distorsionadas, así como daño en la cinta y en la cabeza. Las cabezas deben limpiarse con hisopos de algodón humedecidos en el limpiador de cabezas, haciendo girar los hisopos de algodón contra la superficie de la cabeza y observando la coloración del algodón. Deben emplearse tantos hisopos como sea necesario a fin de evitar su contacto una vez sucios con las superficies de la cabeza. Cuando se observa que el aplicador no se ensucia, la cabeza ya está limpia. Se emplean también hisopos para limpiar las guías de la cinta, los rodillos, el eje motor y el rodillo opresor. Debe tenerse el cuidado especial de limpiar los bordes rebajados de las guías y las ranuras de los rodillos. La acumulación de productos del desgaste en dichas ranuras o en las esquinas de las guías dañará la cinta.

Las cabezas magnetizadas producen resultados impredecibles en los datos reproducidos. *En ninguna circunstancia debe comprobarse la continuidad de una cabeza con un óhmetro.* Una cabeza de grabación magnetizada, sin embargo, produce una alta distorsión de segundo armónico; los efectos desmagnetizadores de la polarización de CA empleados en el proceso de gra-

bación directa reducen la posibilidad de que se magneticen las cabezas de grabación, a menos que se sometan a grandes sobrecorrientes transitorias unidireccionales (como, por ejemplo, las de una comprobación de continuidad) o a la estrecha proximidad a campos magnéticos muy intensos. De este modo, no es prudente retirar las tarjetas electrónicas de la grabadora mientras ésta se encuentra en el modo de grabación. La última sobrecorriente transitoria cuando se retira un amplificador puede magnetizar la cabeza de grabación.

La magnetización de las cabezas de reproducción es más común y afecta su rendimiento de manera impredecible, aunque por lo general el resultado son unos niveles de ruido aumentados. La desmagnetización o desgausamiento de las cabezas de grabación se realiza empleando cualquiera de los limpiadores de cabezas que existen en el mercado. Cuando se limpian las cabezas la grabadora debe estar apagada, y en algunos aparatos el desgausamiento es más fácil y más completo si se retira el ensamblaje de las cabezas. En el manual de instrucciones para el usuario suelen incluirse procedimientos idóneos. Es necesario comprender por completo el mecanismo para retirar las cabezas antes de proceder a quitar el ensamblaje, ya que retirar y volver a colocar este ensamblaje requiere extremados cuidado y atención.

45.2.3 Mecanismos de transporte de la cinta

Desarrollo

El único objetivo del transporte o circulación de la cinta es mover la cinta por las cabezas a velocidad constante y proporcionar los diferentes modos de arrollamiento de operación requeridos para el manejo de la cinta, sin estirarla, distorsionarla o gastarla. A fin de lograr esto, un mecanismo de transporte debe guiar el paso de la cinta por las cabezas con extremada precisión y mantener la tensión apropiada en el área de la cabeza para obtener contacto adecuado entre esta última y la cinta. El bobinado o arrollado de la cinta debe efectuarse suavemente, de modo que en el área de la cabeza se refleje un mínimo de perturbaciones. Debe controlarse el momento de torsión del tensor receptor a fin de obtener un correcto arrollamiento de la cinta en el carrete receptor. También es tarea del mecanismo de transporte mover la cinta de un carrete al otro a gran velocidad en los modos de avance (FF, de *fast-forward*) o rebobinado (*rewind*) rápidos. Incluso con velocidades elevadas, la cinta debe ser manipulada con suavidad y precisión, a fin de que en cada carrete se mantenga un correcto arrollamiento de la cinta. Al pasar del modo rápido a la detención (o viceversa) es necesario mantener control exacto de la cinta, para que ésta no presente flojedad o esfuerzo indebidos. La mayor parte de estas funciones están incluidas en las modernas grabadoras de cinta para instrumentación, aunque constantemente se investigan nuevas posibles mejoras en el área del movimiento uniforme de la cinta.

Las primeras grabadoras para instrumentación fueron un desarrollo ulterior de las mejores grabadoras de audio de la época. En ellas se empleaba el diseño de

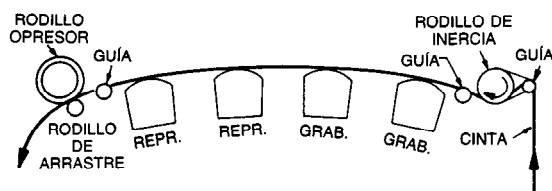


Fig. 45-20. Mecanismo de transporte de ciclo abierto.

mecanismo de transporte de ciclo abierto (fig. 45-20), en el cual la cinta situada en el área de la cabeza tiene sólo un extremo controlado por el eje motor, y el otro extremo es tensado directamente por una función de arrollamiento. A medida que las demandas de instrumentación aumentaron, las máquinas de audio se refinaron y se integraron dispositivos electrónicos especiales; el mecanismo de transporte de la cinta de ciclo abierto predominó hasta la mitad de la década de 1950. Básicamente se trataba de un diseño lógico, sencillo y confiable; sin embargo, su variación de tono dejaba algo que desear.

Existían dos fuentes inherentes de indeseadas variaciones de la velocidad de la cinta en el área de la cabeza. Una provenía de las perturbaciones provocadas por el carrete de suministro, su motor de accionamiento o su mecanismo de freno, y la otra resultaba de vibraciones erráticas en la cinta debido a la gran longitud de cinta no apoyada en el área de la cabeza. La fricción entre la cinta y las cabezas o las guías fijas también inducía vibraciones.

El diseño del mecanismo de transporte de ciclo cerrado que se muestra en la figura 45-21 parecía apropiado para resolver ambos problemas mencionados. La longitud de cinta no apoyada se reduce a la mitad y, si ambos extremos de la cinta en el área de la cabeza se encuentran bajo el control positivo del eje motor, deberían eliminarse las perturbaciones de la función de arrollamiento. En realidad, estas características se logran sólo parcialmente.

Si no existiese deslizamiento o corrimiento entre la cinta y el eje motor en cualquier punto de contacto, podría mantenerse cualquier tensión inicial en el ciclo, pero los únicos elementos capaces de aplicar tensión a la cinta se encuentran fuera del ciclo; es decir, las funciones de arrollamiento. Por tanto, debe existir algo de deslizamiento entre la cinta y el eje motor para que dichas funciones mantengan tensión en el ciclo. Así, incluso en el diseño de mecanismos de transporte de ciclo cerrado, perturbaciones de movimiento en las funciones de arrollamiento siguen provocando variaciones no uniformes en la tensión en el área de la cabeza y variaciones correspondientes en la velocidad de la cinta, pero los efectos son muy reducidos si se comparan con los de los diseños antiguos de ciclo abierto.

Hacia 1958 casi todos los fabricantes utilizaban mecanismos de transporte de ciclo cerrado, pero aún se esforzaban por obtener mejor rendimiento. Por esa época un fabricante desarrolló un nuevo diseño de accionamiento del eje motor que proporciona acción diferencial. Básicamente, se trata de un eje motor con

dos diámetros discretos (con una diferencia de sólo algunas milésimas de pulgada). Como se muestra en la figura 45-22, se utilizan rodillos opresores de contorno especial, y la cinta que entra en el área de la cabeza es forzada contra el diámetro menor del eje, mientras que la cinta que abandona el área de la cabeza es forzada contra el diámetro mayor. Así se genera una tensión en el ciclo debida a la diferencia de velocidad entre la cinta que entra y la cinta que abandona el ciclo. En este diseño se requiere menor aporte de tensión por las funciones de arrollamiento, y el área de la cabeza está más aislada de las perturbaciones de dichas funciones. Sin embargo, tal accionamiento no es adecuado para operación bidireccional.

Actualmente se emplea mucho otro tipo de configuración del mecanismo de transporte, de dos ejes motores, que se muestra en la figura 45-23, denominado sistema de transporte de doble eje motor o de ejes motores diferenciales. En algunos diseños ambos ejes giran a la misma velocidad periférica y las funciones de arrollamiento son la base para establecer la tensión de la cinta en el área de la cabeza, de manera bastante parecida a como ocurre en el caso del mecanismo de transporte de ciclo cerrado ordinario. Sin embargo, en la mayor parte de los aparatos con dos ejes motores éstos giran con velocidades periféricas ligeramente diferentes y establecen la tensión de la cinta en el área de la cabeza con esta acción diferencial. De esta forma se requiere menos tensión externa al área de la cabeza, y se obtiene considerable aislamiento respecto a las perturbaciones de la función del arrollamiento.

A mediados de la década de 1960 reapareció el mecanismo de transporte de ciclo abierto en una forma refinada que compite bastante bien con los diseños más elaborados de ciclo cerrado. Un esfuerzo considerable en el diseño de un fabricante, que se muestra en la figura 45-24, estaba condenado al uso de brazos amortiguadores giratorios con un eje motor de alta inercia y elevado momento de torsión. La significativa reducción en la longitud de la cinta no apoyada permitió una mejora en la vibración de tono de alta frecuencia.

Con todo, los nuevos diseños de ciclo abierto preservan la sencillez de los antiguos y proporcionan un rendimiento bastante mejorado a costo considerablemente inferior que la mayor parte de los aparatos de ciclo cerrado.

Otra característica que hizo su aparición en los diseños de sistemas de transporte es la eliminación de los rodillos opresores. En estos sistemas de accionamiento

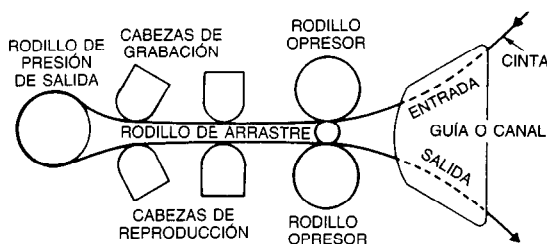


Fig. 45-21. Mecanismo de transporte de ciclo cerrado.

se utiliza un eje motor con elevada fricción superficial y grandes ángulos de envolvente (contacto) en la cinta. Cuando se emplean con el accionamiento de doble eje motor (ejes diferenciales), se obtiene un razonable aislamiento respecto a las perturbaciones del arrollamiento. Sin embargo, un diseño es parecido al del sistema de transporte de ciclo cerrado, con el rodillo de vuelta completa sustituido por un eje motor de alta fricción superficial. En este aparato, la tensión de la cinta en el área de la cabeza y el aislamiento contra las perturbaciones del arrollamiento se obtienen por medio de canales para cinta controlados al vacío antes y después de las cabezas de grabación y reproducción. Se dice que el rendimiento de este aparato en cuanto a vibración de tono y sesgo es excelente.

Con la eliminación del rodillo opresor se elimina una fuente más de irregularidad en la velocidad y el guiado de la cinta. Los rodillos opresores con deformaciones en la superficie, puntos resbaladizos o pegajosos o cojinetes ruidosos pueden ser contribuyentes importantes a la vibración de tono y el sesgo dinámico de un aparato.

Mecanismos de arrollamiento de cinta

Los primeros mecanismos de transporte de cinta cumplían las funciones de arrollamiento mediante la sencilla colocación de un motor de par tanto en el eje del carrete de suministro como en el del carrete receptor. Se han utilizado motores de par de CA y de CD, y están diseñados para proporcionar un momento de torsión esencialmente constante en un amplio intervalo de velocidades sin sobrecalentarse a velocidades extremadamente bajas. La tensión de retención en el lado del carrete de suministro se obtiene mediante un sencillo freno de fricción o por medio de un momento de torsión eléctrico invertido. Sin embargo, la tensión creada de esta forma varía del carrete lleno al carrete vacío, debido a los radios cambiantes del rollo formado por la cinta. La tensión mínima se obtiene con un carrete lleno, y la máxima, con uno casi vacío. Ambos esquemas también contribuyen a variaciones de tensión a corto plazo debido a una acción irregular de frenado o pulsaciones intermitentes en el motor de torsión. Cuando tales variaciones de tensión se transmiten al área de la cabeza resulta afectado el rendimiento de vibración de tono del mecanismo de transporte. Se han dedicado muchos esfuerzos de diseño a eliminar estas perturbaciones, con diferentes grados de éxito. En la generalidad de los esquemas se hace uso de alguna forma de detector del tamaño (radio) del rollo formado por la cinta a fin de determinar la cantidad de cinta en el carrete. En un método se emplea un brazo con un rodillo en el extremo que se desplaza sobre el rollo de cinta. A medida que se mueve el brazo, un varillaje controla la acción de un freno de fricción mecánico. En una versión parecida se emplea el mismo brazo, a fin de regular la posición de un potenciómetro que controla el voltaje tensor aplicado al motor de torsión. Otros dispositivos detectores son unas celdas fotoeléctricas y un muelle ligeramente cargado que sigue el punto de salida de la cinta del carrete. Por supuesto, la necesidad de mecanismos de control tan elaborados depende

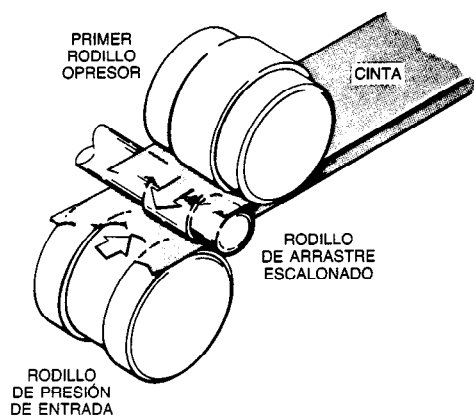


Fig. 45-22. Eje motor de dos diámetros.

del diseño del mecanismo de transporte; sólo en los aparatos más vulnerables a perturbaciones de bobinado es necesario utilizar estas técnicas.

Existe una característica altamente deseable que probablemente deba incluirse en todo mecanismo de bobinado de grabadora de cinta. Ésta es una función contra fallos en forma de freno de banda o de disco montado en los ejes del carrete a fin de llevar la cinta a una detención controlada en caso de un fallo de energía. Algunos aparatos antiguos no tienen esta característica, y en los modos de avance o de rebobinado rápidos ese fallo de energía puede colocar varios cientos de metros de datos irremplazables en un horrible montón de cinta arrugada, distendida y completamente arruinada.

El empaquetamiento de las grabadoras de cinta está determinado en alguna medida por su uso. Las aplicaciones han impuesto serios límites al tamaño de la grabadora, pero el usuario no ha estado dispuesto a reducir el tiempo de grabación. Como resultado, se

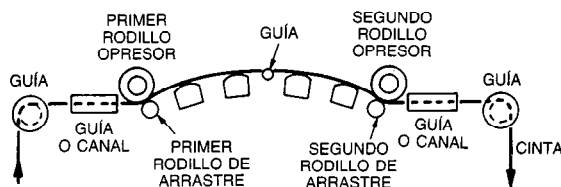


Fig. 45-23. Mecanismo de transporte de doble eje motor de ciclo cerrado.

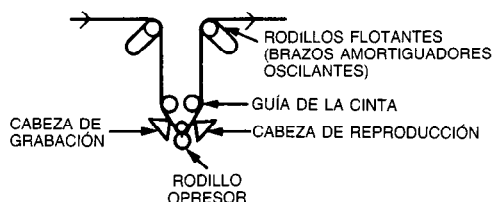


Fig. 45-24. Mecanismo de transporte de ciclo abierto refinado.

han desarrollado varias configuraciones de carretes, con frecuencia a expensas de la sencillez de la grabadora y facilidad de arrollamiento y carga-descarga de la cinta.

El mecanismo de transporte de la cinta con carretes uno al lado del otro, o sea, en el que ambos carretes giran en el mismo plano, es el más conveniente y convencional para hacer que la cinta pase por las cabezas. Sin embargo, las grabadoras espaciales y otros tipos de grabadoras para adquisición de datos, en las que el volumen es crítico, utilizan un mecanismo de transporte de la cinta con carretes concéntricos, en el cual los carretes se apilan uno encima del otro. Las grabadoras de carretes concéntricos requieren un ensamblaje mecánico complicado en el cubo del carrete, tanto en lo que se refiere a los mecanismos de accionamiento como de carga y descarga de los carretes. También cuentan con una trayectoria complicada de la cinta y, a menos que se tenga bastante cuidado al diseñarlas, tendrán deficientes características de sesgo.

Sistemas de control de la velocidad

Dado que se supone que el eje motor controla la velocidad de la cinta, su potencia de accionamiento en las primeras grabadoras de cinta se eligió para velocidad constante. La elección más popular fue el motor sincrónico de histéresis. Tal motor marcha sincronizado (enclavado) en fase con la frecuencia de la alimentación de CA, por lo que su estabilidad de velocidad a largo plazo es tan buena como la estabilidad de frecuencia de la línea de alimentación si se opera con carga constante. Con carga variable variará su relación de fase con la frecuencia de alimentación, pero no se deslizará un polo a menos que se sobrecargue. La construcción de polos de estos motores también provoca una pequeña variación periódica o intermitencia en su salida de momento de torsión. Debe proporcionarse considerable alisamiento por medio de un volante y algunas veces mediante la elasticidad de una banda impulsora. Tal sistema se describe mejor como uno de accionamiento de gran inercia y bajo momento de torsión.

La dependencia respecto a la frecuencia de la línea de alimentación para fines de estabilidad es una desventaja importante en muchas aplicaciones. Incluso en aplicaciones de laboratorio, con una fuente de alimentación pública de energía puede esperarse alguna inestabilidad. Tal fuente de energía tendrá excelente estabilidad a largo plazo, pero no son extrañas ciertas variaciones a corto plazo hasta de $\pm 0.25\%$. Hace mucho, estos factores condujeron al desarrollo de fuentes de energía de precisión de 60 Hz consistentes en fuentes de frecuencia muy estables de 60 Hz (diapasones u osciladores de cristal de cuenta regresiva) y amplificadores de potencia con capacidad de 50 a 150 W. Los ejes motores accionados de esta forma son independientes de las fluctuaciones de frecuencia de la línea de alimentación, pero adolecen de desventajas en el tamaño, peso y consumo de energía.

Se han desarrollado algunas técnicas bastante satisfactorias de accionamiento a velocidad constante, utilizando motores de CD. En uno de los primeros mé-

todos se colocaba un generador tacométrico de CA sobre el eje motor de transporte. Su salida era rectificada y comparada con un voltaje CD de referencia. A continuación, el voltaje del comparador se utilizaba para controlar la velocidad del motor del rodillo de transporte.

En otra técnica se acciona un generador de tono con el motor del eje de transporte y su salida se dirige a una serie de divisores de frecuencia. Después, un conmutador selector de velocidad elige la salida apropiada del divisor para fines de comparación de fase con una frecuencia de referencia altamente estable. La salida del comparador de fase se utiliza para controlar la corriente de campo del motor del eje de transporte de CD devanado en derivación. Cuando ambas señales se enclavan en fase, la velocidad del motor es constante. Este método de control, además de ser independiente de la frecuencia de la línea, presenta varias ventajas sobre el accionamiento de CA. Una de ellas se relaciona con la facilidad para cambiar las velocidades de la cinta. La simple conmutación de uno más de los circuitos de "división entre dos" en el circuito divisor duplicará la velocidad del motor. Otra ventaja es la reducción en tamaño, peso y consumo de potencia, ya que no se requiere un amplificador de alta potencia y baja eficacia.

Los primeros ejes motores de transporte accionados por CD también eran sistemas con una inercia relativamente alta, aunque con algo más de momento torsional que los mecanismos de CA. Cada vez se usan más los magnetomotores de CD (de imán permanente) en los que se emplea un rotor de circuito impreso con muchos polos. Este dispositivo ha hecho posible un sistema de accionamiento de baja inercia y alto momento de torsión, y varios fabricantes lo han incorporado a sus grabadoras.

Existen velocidades estándar de cinta de 120, 60, 30, 15, 7 1/2, 3 3/4, 1 7/8, 15/16 y 15/32 pulg/s, y la mayor parte de las grabadoras para instrumentación cuentan con un sistema de fácil elección de tres a seis de ellas. Algunas de las primeras grabadoras requerían cambios de banda impulsora si se necesitaba un cambio de velocidad mayor que 2:1.

Cualquiera de estos métodos de control para el accionamiento del eje motor de transporte proporciona una velocidad de cinta razonablemente exacta en el modo de grabación. Sin embargo, si se requiere la reproducción precisa de los datos grabados, en el modo de reproducción es necesario un servomecanismo de control de la velocidad. Para esto se emplean varias técnicas; todas ellas operan a partir de una señal de referencia grabada en la cinta con las señales de datos.

Si el aparato de grabación es accionado por un motor de eje de transporte sincrónico de histéresis, su fuente de precisión de 60 Hz se utiliza como la señal de referencia y se graba mediante la modulación de una portadora de 17 kHz. Si un motor del eje de transporte de CD (diferente de las nuevas versiones de baja inercia) acciona la grabadora, la referencia de 60 Hz se genera a partir del oscilador de referencia que determina la velocidad y se graba de la misma forma. Por tanto, cualquiera de los dos sistemas es capaz de pro-

porcionar una señal de referencia que puede ser empleada por cualquiera de estos sistemas. El servocontrol de un aparato de reproducción sincrónico de histéresis se logra al recuperar la señal de referencia reproducida de 60 Hz y comparándola en fase con el voltaje de referencia local de precisión de 60 Hz. La salida del comparador es una señal de CD o una de CA de muy baja frecuencia, que controla la salida de un oscilador de 60 Hz controlado por voltaje. Esta salida es ampliada y usada para accionar el motor del eje de transporte. La acción del servomecanismo enclava o sincroniza de manera efectiva la señal de referencia grabada con la señal de referencia local de precisión, y para cada ciclo de la referencia local se hace pasar por el aparato una longitud de cinta equivalente a un ciclo de la referencia grabada. El servocontrol de un aparato de reproducción con accionamiento del eje de transporte por CD suele efectuarse sustituyendo la señal local generadora de tono por la señal de referencia grabada reproducida de la cinta.

Ambos métodos son sistemas de comparación de fase, por lo que básicamente son servomecanismos posicionales. El intervalo de sincronización está limitado por la gran masa de los impulsores del eje del rodillo de transporte, pero la mayor capacidad de momento torsional del motor de CD proporciona a ese sistema un pequeño margen. Tal vez el límite de su respuesta sean intervalos de sincronización de 1/2 a 3 o 4 ciclos.

La aparición de los motores CD de baja inercia y alto momento torsional para ser empleados en impulsores de ejes de transporte conllevó la existencia de un control totalmente diferente de la velocidad de reproducción. La rápida respuesta de estos dispositivos hizo posible eliminar prácticamente todas las variaciones de velocidad a corto y largo plazos hasta 100 Hz o más. En la mayor parte de estos sistemas se emplea un doble servomecanismo de control que abarca una función de control de velocidad y una función de control posicional. Para grabación existe un generador tacométrico de CA conectado al eje del transporte que reproduce varios cientos de ciclos para cada rotación del eje citado. Las variaciones en la frecuencia de esta señal tacométrica son detectadas por un discriminador y se convierten en un voltaje de control de CD para el motor del eje de transporte. Este control, que básicamente es un servomecanismo de velocidad, se utiliza a fin de hacer que el eje de transporte gire a la velocidad apropiada. La fase de la señal tacométrica también se compara con un oscilador de referencia local una vez que se ha alcanzado la velocidad apropiada y se ha establecido la operación de enclavamiento de fase. Durante la grabación, la señal del oscilador de referencia se graba también en la cinta. Para el servocontrol de velocidad de la cinta en el modo de reproducción, la referencia grabada se emplea para sustituir el generador tacométrico, y se realiza esencialmente la misma operación que se acaba de describir.

Las especificaciones para la vibración de tono son bastante parecidas tanto para el sistema de accionamiento del eje de transporte de alta inercia como para el de baja inercia, en lo que a vibración porcentual de tono corresponde. Sin embargo, existe una

notoria diferencia en las componentes espectrales de la vibración de tono en los dos sistemas. El sistema de alta inercia presenta mayores amplitudes de frecuencias de baja vibración de tono, y sólo la vibración acostumbrada de mayor frecuencia provocada por rayaduras y vibración de la cinta. Cuando los sistemas de baja inercia están servocontrolados a partir de la cinta, prácticamente eliminan la vibración de tono de baja frecuencia, pero, al ser tan sensibles, en realidad contribuyen a la vibración de tono de alta frecuencia normal en la forma de dispersión espectral.

Aun cuando parece que ambos sistemas presentan aproximadamente el mismo porcentaje de vibración de tono total, existe un punto importante que favorece al sistema de baja inercia: la enorme mejora en el error de la base de tiempo (TBE). El TBE especificado en algunas máquinas es $\pm 0.5\%$ μs absoluto. Esto significa que entre dos puntos cualesquiera de la cinta el error de sincronización no será mayor de $\pm 0.5\ \mu s \pm$ la tolerancia de referencia del cristal.

Irregularidades en los movimientos de la cinta

El movimiento ideal de la cinta puede definirse sencillamente. La cinta debe moverse sobre las cabezas con velocidad absolutamente uniforme, conocida con precisión. En realidad, ningún mecanismo de transporte de cinta logrará jamás este movimiento ideal, aunque continuamente se logran mejoras. Las desviaciones a mediano y largo plazo respecto de la velocidad promedio deseada pueden corregirse mediante servomecanismos, como ya se analizó antes. De hecho, los ejes de transporte de baja inercia y alto momento torsional son lo suficientemente rápidos para corregir algunas de las variaciones a corto plazo, pero, en general, de esta manera no es posible eliminarlas todas. Como resultado, deben considerarse entre las características fundamentales de una grabadora.

Vibración de tono. En un mecanismo de transporte de cinta es posible provocar de muchas formas variaciones de velocidad a corto plazo uniformes de un extremo a otro de la cinta. Algunas de ellas son pulsaciones de los motores de par, excentricidades de los carretes, irregularidades en el arrollamiento de la cinta o en las características físicas de ésta, vibraciones en la cinta provocadas por el rozamiento a medida que pasa sobre guías o cabezas fijas, desgaste mecánico de partes giratorias, ligeras intermitencias del motor de accionamiento del eje de transporte, oscilaciones momentáneas (transitorias) de voltaje en la línea de alimentación que podrían afectar los motores, rodillos opresores con deformaciones en la superficie y cojinetes pegajosos. El problema se complica más por el hecho de que los carretes y los ensamblajes de accionamiento de los carretes tienen velocidades variables y una masa que varía constantemente.

Las variaciones de velocidad que son uniformes de un extremo a otro de la cinta han sido descritas como vibración de tono y deriva o desplazamiento. La vibración de tono denota variaciones en la velocidad que ocurren a frecuencias superiores a 0.10 Hz, y la deriva (o exactitud de la velocidad de la cinta) se emplea para

las frecuencias menores de 0.1 Hz. Según se aplica a las grabadoras para instrumentación, el uso común ha generalizado la definición del término "vibración de tono" a fin de incluir todas las variaciones de 0.1 Hz a 10 kHz. "Error por desplazamiento temporal" (TDE, de *time displacement error*), "error de la base de tiempo" (TBE, de *time base error*) y "agitación" ("jitter") son términos utilizados para describir desde diferentes puntos de vista las mismas variaciones de velocidad de la cinta. Algunas veces estos términos se emplean de forma inapropiada, por lo que pueden inducir confusión al tratar de describir la capacidad de una grabadora para reproducir una señal con sus relaciones de tiempo originales. Debe recordarse que las cifras de TDE, TBE y agitación (o fluctuación) deben establecer el tiempo durante el cual fueron medidas, y la vibración de tono debe citarse ya sea en valores rms (eficaces) o en valores de cresta a cresta sobre una banda específica de frecuencias.

El espectro de vibración de tono de un aparato bien diseñado está constituido por una combinación de pequeñas componentes sinusoidales discretas y una señal de ruido distribuida más o menos uniformemente. Debido al ruido implicado, es difícil efectuar una medición clara y bien definida de la vibración de tono. La señal de la vibración de tono en sí puede obtenerse grabando una onda sinusoidal de referencia extremadamente estable y haciendo pasar la señal reproducida a través de un discriminador de FM. La señal generada de esta forma sería cero si la velocidad de la cinta fuese exactamente la misma para reproducción que para grabación. Cualquier variación entre las dos velocidades desvía la frecuencia del tono de referencia y produce una salida del discriminador. Se acostumbra medir las componentes de la vibración de tono hasta de, por lo menos, 10 kHz en grabadoras para instrumentación. El medidor de vibración de tono es el dispositivo de mayor uso para efectuar mediciones cresta a cresta de la vibración de tono cuando se comprueban las especificaciones de dicha vibración.

Una forma común de especificación de vibración de tono es la medición de "vibración de tono acumulativa", que puede realizarse haciendo pasar la señal de vibración de tono a través de un filtro pasabajas de corte variable y midiendo la salida del filtro para valores crecientes de la frecuencia de corte.

Dado que la contribución del ruido a la señal de vibración de tono es esencialmente uniforme, el perfil de la curva acumulativa aumenta con la frecuencia. En cada punto en el que alguna componente giratoria genera una contribución sinusoidal discreta existe una pequeña función escalonada en la curva acumulativa. Muchos fabricantes publican curvas de este tipo, pero por lo general son resultados promedio de pruebas de muchos mecanismos de transporte y no presentan los casos extremos que es posible encontrar en aparatos individuales.

Para la vibración de tono se ha utilizado durante mucho tiempo un valor rms en el caso de los aparatos de audio, pero la medición de cresta a cresta es más útil para las grabadoras usadas en instrumentación. En realidad, es casi imposible lograr un valor rms (eficaz)

verdadero de la vibración de tono, ya que la señal de la vibración contiene una componente de CD, una componente de ruido y muchas componentes de ondas sinusoidales. Como aproximación bastante burda, puede suponerse que el valor rms (eficaz) es de 1/6 a 1/4 del valor de cresta a cresta.

Para medir las perturbaciones de la base de tiempo provocadas por la vibración de tono se emplean muchas técnicas; todas implican la comparación entre un retardo electrónico preciso y la base de tiempo representada por alguna longitud de la cinta. Se utiliza un pulso reproducido de la cinta para iniciar el retardo electrónico, y algún periodo más tarde se compara un segundo pulso de la cinta con el pulso retardo electrónicamente, a fin de comprobar la coincidencia temporal. Entonces la discordancia temporal entre los dos pulsos representa un error de base de tiempo atribuible a variaciones en la velocidad de la cinta. El equipo para efectuar tales mediciones con exactitud es costoso y difícil de encontrar en el mercado.

Sesgo. El término "sesgo" se emplea para describir las diferencias temporales fijas y variables entre las diversas pistas de un solo apilamiento de cabezas. Implica que la cinta se mueve de forma no longitudinal a medida que pasa por las cabezas; es decir, que se está sesgando o que está guiñando. El sesgo fijo, o estático, contribuye a una diferencia de sincronización relativa constante; el sesgo dinámico produce una diferencia de sincronización variable entre pistas.

El sesgo fijo suele ser provocado por desalineamiento de la cabeza con relación a la cinta, desalineamiento de las cabezas individuales en un apilamiento de cabezas, y paso incorrecto de la cinta por las guías, lo que produce diferencias fijas en la distribución de la tensión a medida que la cinta cruza por las cabezas. El sesgo dinámico se produce cuando existe una distribución no uniforme de la tensión en la cinta. Los productores de vibración de tono, tales como rayaduras y vibración de la cinta, iniciarán sesgo dinámico.

El sesgo fijo y algunas formas de sesgo dinámico producen errores de sincronización relativa entre las pistas, que son proporcionales a la separación entre éstas. Sin embargo, ciertas irregularidades dimensionales de la cinta y componentes aleatorios de la vibración de tono provocan algo de sesgo dinámico, y los errores de sincronización producidos de esta forma no están correlacionados.

Como en el caso de la vibración de tono, los errores de sesgo son impuestos tanto por la operación de grabación como por la de reproducción. Las cintas reproducidas en el mismo aparato en que fueron grabadas presentarán menor error de sesgo que las reproducidas en aparatos diferentes.

En las especificaciones de los mecanismos de transporte de la cinta suelen incluirse el sesgo, el sesgo dinámico o el error de la base de tiempo intercanales total (estático y dinámico). Se proporcionan valores para diferentes velocidades de la cinta, y algunas de las máquinas modernas son capaces de proporcionar $\pm 0.5\%$ μ s o incluso una cifra mejor a 120 pulg/s (ips) entre pistas adyacentes en la misma pila de cabezas.

Alargamiento. Otro tipo de error de sincronización ocurre entre pistas con numeración par e impar, debido a la separación entre las dos pilas de cabezas. Si las dimensiones físicas de la cinta varían debido al medio ambiente o se alargan (estiran) después de haber hecho la grabación, un error correspondiente de sincronización ocurrirá durante la reproducción. Como ejemplo exagerado, supóngase que habrá un cambio de un 1 % en alguna parte de la longitud de la cinta. Esto representa 15 mpulg para una distancia de 1.5 pulg, y equivaldría a muchas longitudes de onda de la señal grabada. Conforme menor sea la distancia de separación entre las pilas de cabezas, menor será la pérdida de señal observada para el mismo cambio en la longitud de la cinta. Las diferencias en la tensión de la cinta entre los aparatos de grabación y de reproducción también pueden contribuir a este tipo de error de sincronización.

Se producirán grandes errores de sincronización si los datos se graban en un aparato y se reproducen en otro cuyo formato de pistas sea diferente.

Efectos de la vibración de tono. La vibración de tono tiene efectos dañinos sobre los datos grabados. Tal vez el más conocido sea el ruido que produce en las grabaciones por FM con portadoras. Una portadora de frecuencia constante será modulada en frecuencia por variaciones en la velocidad de la cinta; cuando la portadora se reproduce y discrimina, la salida esperada de CD también incluye el ruido de la modulación indeseada. Cuando la portadora es modulada con una señal de datos, este ruido se agrega a los datos en el proceso de desmodulación. Por tanto, el efecto de primer orden de la vibración de tono es incrementar el nivel de ruido de la señal reproducida con una reducción correspondiente en el intervalo dinámico. Para una vibración porcentual de tono dada, la capacidad del ruido de la vibración de tono de interferir en la señal de los datos depende de la razón de desviación del sistema de FM con portadora que se esté grabando.

En el modo de grabación directa, la vibración de tono perturba en primer lugar la base de tiempo de la señal reproducida. De manera parecida, la onda de forma reproducida presentará un ensanchamiento de sus componentes espectrales en un análisis espectral.

45.2.4 Circuitos electrónicos

Los circuitos electrónicos de una grabadora magnética para instrumentación realizan dos tareas diferentes: controlar la velocidad de la cinta y manejar los datos o acondicionar la señal. Cada una se efectúa de forma independiente de la otra; la interdependencia se da en la calidad de los datos reproducidos. En la sección anterior se analizaron los sistemas de control de la velocidad de la cinta, por lo que aquí se tratarán los circuitos electrónicos restantes. Cada tipo de operación de grabación-reproducción requiere circuitos propios más bien especiales, pero sólo se describen los modos directo y de FM con portadora única. Con los circuitos electrónicos de FM se presentarán brevemente los circuitos de compensación de la vibración de tono.

Grabación y reproducción directas

Las operaciones de grabación y reproducción directas requieren una cabeza de amplificación, un oscilador de polarización, un preamplificador de reproducción y un amplificador-ecualizador de reproducción. La cabeza de amplificación se denomina más comúnmente amplificador de grabación y realiza varias funciones. Debe presentar una impedancia nominal alta a la señal de los datos, a fin de minimizar la carga de la fuente de la señal. Una primera aproximación de sus características ganancia/frecuencia sugeriría una salida de corriente constante para todas las frecuencias. Por desgracia, la impedancia de la cabeza de grabación cambiará drásticamente en el intervalo de frecuencias cubierto por los aparatos modernos de banda ancha. Además, como ya se dijo, la grabación con flujo constante (igual intensidad magnética en la cinta para todas las frecuencias) es un caso ideal, ya que las pérdidas en cabeza y la desmagnetización de la grabación son funciones de la frecuencia.

Por estas razones se acostumbra a que el amplificador de grabación proporcione una salida aumentada, o preénfasis, a las frecuencias superiores. No debe confundirse esto con el preénfasis exagerado que se utiliza en los aparatos de audio: sólo se emplea el refuerzo de alta frecuencia requerido para aproximarse a una grabación con flujo constante. Por ejemplo, un aparato de banda ancha proporciona de 8 a 10 dB de preénfasis a 1.5 MHz. Otra función del amplificador de grabación es agregar la señal polarizada a la señal de los datos al mismo tiempo que constituye un separador para el oscilador de polarización.

En alguna parte de los circuitos electrónicos para grabación suele incluirse un punto de control, a fin de observar la onda de forma de la corriente de la cabeza de grabación. Por lo general, el nivel de control es aproximadamente 10Ω mayor que el de tierra y presenta una señal de muy bajo nivel. Para una visión idónea se requiere considerable ganancia en el osciloscopio. Las ondas de forma en este punto, especialmente en los aparatos de banda ancha, pueden presentarse bastante distorsionadas, por lo que es necesario consultar en el manual de instrucciones del aparato los detalles de su empleo. Los niveles de grabación deben establecerse de modo que se obtenga una equivalencia óptima entre los intervalos dinámicos de la señal de datos y la grabadora. Un amplificador de grabación bien diseñado no se saturará antes que la cinta, y debe recordarse que la saturación de la cinta es de alguna forma un proceso gradual, con recuperación muy rápida. La distorsión que resulta cuando la cinta se sobregaba puede tolerarse algunas veces a cambio de un ligero aumento del intervalo dinámico.

Con frecuencia, cada amplificador de grabación cuenta con un medidor del nivel de grabación, pero se emplea poco si la señal de datos no es sinusoidal. Los trenes de pulsos de bajo factor de trabajo pueden saturar drásticamente la cinta con poca o ninguna indicación en el medidor rms (eficaz) ordinario. Algunos fabricantes han ofrecido como accesorio medidores de valores de cresta, pero por desgracia no son de uso común.

El oscilador de polarización debe operar a varias veces (3.5 a 5 o más) la frecuencia más alta que pueda manejar la grabadora. Se requiere una onda de forma sinusoidal pura, ya que cualquier distorsión será reflejada por la señal grabada. A fin de eliminar los problemas de sincronización o de "batido" de frecuencias se utiliza un oscilador de polarización única.

Un amplificador separador, por lo general localizado en el amplificador de grabación, proporciona el accionamiento de polarización necesario para cada pista y a la vez elimina la posibilidad de que ocurra interferencia entre los canales a través del cableado de distribución de la polarización. Algunas veces se incluyen, integrados a este amplificador separador, medios convenientes para monitorizar la corriente de polarización en el medidor del nivel de grabación.

La salida de una cabeza de reproducción puede ser una señal muy pequeña —es decir, de unos cuantos microvolts— en un aparato de banda ancha. A estos niveles, la señal es sometida a toda clase de captación de ruido y problemas de ciclos conectados a tierra. La salida de la cabeza de reproducción se aplica de manera inmediata a un preamplificador, a través de un cable cuidadosamente blindado de longitud mínima. En gran medida, este preamplificador establece el nivel de ruido del sistema, por lo que es necesario tener gran cuidado en su diseño a fin de proporcionar una operación a bajo ruido y óptimo acoplamiento de impedancia con la cabeza. El diseño ideal probablemente es el que resulta con preamplificadores de circuitos integrados incluidos en la pila de cabezas de reproducción.

La siguiente operación en la señal reproducida se realiza con un amplificador de reproducción. Ésta es la unidad que ecualiza la característica salida/frecuencia de la cabeza de reproducción. Cuando se emplea un amplificador de voltaje, es necesario insertar una atenuación progresiva de 6 dB por octava desde la frecuencia más baja hasta el pico de la frecuencia central en la curva de salida de la cabeza. En este punto, sin importar el tipo de amplificador, es necesario agregar una cantidad creciente de ganancia (hasta de 12 o 18 dB por octava) a fin de compensar la característica en decremento de la salida de la cabeza y, en alguna medida, otras pérdidas de alta frecuencia. Para cada velocidad de la cinta se requiere un circuito diferente de ecualización a fin de compensar apropiadamente las distintas curvas de salida de la cabeza, y cada circuito cuenta con dos a cuatro ajustes asociados. Cada ajuste afecta una porción específica de la curva de respuesta, y cuando se establece apropiadamente puede proporcionar la respuesta especificada del aparato. En especial en los aparatos de banda ancha, uno de los ajustes suele asociarse con la respuesta de fase. Se emplea para proporcionar ecualización de fase o, en otras palabras, para asegurar que las diversas componentes espectrales lleguen a la salida con la relación temporal idónea.

La ecualización de fase es especialmente aconsejable cuando se reproducen señales de datos de tipo pulsátil, a fin de minimizar la sobrecorrección en la onda de forma reproducida.

La elección de la red (circuito) de ecualización adecuada para la velocidad de la cinta varía de un tipo de aparato a otro. En algunos, la elección es automática con la selección de la velocidad; en otras, se requiere conmutación en un control de panel frontal; otras requieren la recolocación de unidades (circuitos) enchufables cuando se cambian las velocidades.

La última función del amplificador de reproducción es llevar la señal a un nivel estándar de impedancia y voltaje de salida, y tal vez proporcionar conmutación a un medidor para fines de monitorización.

Grabación y reproducción por FM

Los circuitos electrónicos para grabación por FM con portadora única son proporcionados por el fabricante de la grabadora en forma de unidades enchufables que es posible intercambiar con las unidades de grabación y reproducción directas. Un amplificador de grabación por FM contiene un amplificador, un oscilador cuya frecuencia varía con la señal de datos y un amplificador de salida empleado esencialmente como acoplador de impedancia para la cabeza de grabación. Como ya se indicó, la señal se graba hasta la saturación en la cinta a amplitud constante. Cada velocidad de la cinta requiere una frecuencia central diferente; en el amplificador de grabación esto se logra mediante unidades enchufables o eléctricamente conmutables determinadoras de la frecuencia. Los circuitos típicos del oscilador son multivibradores o un diseño fantastrón modificado.

Un amplificador de reproducción por FM contiene un amplificador para elevar el nivel de la salida de la cabeza, un limitador, un desmodulador de FM, un filtro pasabajas y un amplificador de salida a fin de llevar la señal a un nivel estándar de impedancia y voltaje de salida. El diseño del primer amplificador no es crítico, porque la señal saturada ya presenta una excelente relación señal-ruido en la cinta, y la frecuencia central está ubicada a fin de minimizar variaciones en la salida de la cabeza sobre las frecuencias utilizadas. Para eliminar las variaciones en la amplitud y proporcionar una señal estandarizada al desmodulador se emplea una operación de limitación estricta. Se usan varios tipos de desmodulador. El método más sencillo no es más que un detector de cruce por cero, que activa un impulso de energía constante para cada cruce. Puede utilizarse un multivibrador monoestable; el verdadero desafío es el desarrollo del pulso de energía constante. Tanto la duración como la amplitud del impulso deben ser precisamente constantes, ya que constituyen la base para la exactitud de la unidad de reproducción. El tren de pulsos resultante se hace pasar a continuación a través de un filtro pasabajas a fin de eliminar las componentes de ruido y residuales de la portadora, y se amplifica para poder proporcionar una salida estandarizada.

Otra técnica de desmodulación que se emplea algunas veces consiste en un oscilador de fase sincronizada cuya frecuencia se compara respecto a la fase con la señal reproducida. Las diferencias en las dos frecuencias producen una salida que se emplea para accionar el oscilador local a fin de obtener un valor cero. Por

tanto, el voltaje de excitación o accionamiento es un duplicado de la señal de datos. De nuevo se utiliza un filtro pasabajas en la salida. Como en el caso del amplificador de grabación por FM, para establecer las frecuencias centrales y las características del filtro se emplean unidades enchufables o elementos eléctricamente conmutables.

Los efectos dañinos de la vibración de tono provocan grave deterioro en una señal grabada por FM. Si deben obtenerse las más altas resoluciones es necesario utilizar un método de compensación de la vibración de tono. Una de las formas más sencillas de compensación se obtiene por medio de la desmodulación de una señal de referencia que fue grabada con los datos (en una pista diferente), generando así la señal de vibración de tono en sí. Después, sencillamente se resta esta señal de los canales de muchos datos, a medida que son desmodulados.

Debe tenerse bastante cuidado para compensar las características de fase y ganancia tanto del desmodulador de referencia como del de datos. En un sistema parecido la señal se recupera en forma semejante de la referencia grabada, pero se emplea para controlar la duración de los pulsos de energía constante que se generan en el desmodulador de datos.

En otras técnicas de grabación por FM se emplean FM de banda ancha, subportadora de FM de ancho de banda constante, y subportadora IRIG de ancho de banda proporcional. El equipo para utilizar estas técnicas es proporcionado casi por completo por los fabricantes de equipo terrestre de telemetría, y todas sus señales son grabadas mediante el modo directo de grabación. No son peculiares de las grabadoras de cinta, y no se analizarán aquí.

Otras características

Control remoto. Casi todos los modos de operación de las grabadoras magnéticas están controlados por relevadores o conmutadores electrónicos. Cuando se emplea control remoto, el modo de operación es controlado por conmutadores de contacto momentáneo, ya sea cerrando o abriendo, en el caso del circuito de detención, los circuitos apropiados. El control remoto permite la activación de todos los modos de los mecanismos de transporte a partir de otra posición, excepto las de encendido y apagado. Otras indicaciones del modo de operación de las grabadoras que con frecuencia se presentan en el control remoto son un indicador de rotura de la cinta, un medidor de la cinta que queda por correr, una indicación de la velocidad de la cinta, y un servomecanismo indicador de la sincronización del eje transportador. Con la atención actual a las normas de interconexión entre los dispositivos electrónicos de medición, la posibilidad de interconectar las capacidades remotas de las grabadoras para instrumentación incrementa aún más la flexibilidad del aparato.

Grabación bidireccional. Esta posibilidad se ha vuelto una realidad gracias a los mecanismos de transporte de doble eje motor y los servosistemas de arrollamiento

mejorados. Por lo menos un fabricante ha empleado esta técnica para ampliar en gran medida el tiempo de grabación de algunos canales de datos. Si sólo se van a grabar 2, 4 o 7 señales de datos, es posible utilizar un circuito especial de seriación de la siguiente manera. Se pone en operación la grabadora con el número deseado de pistas por grabar. Al término de cada paso, la grabadora detecta el extremo de cinta utilizable, invierte la dirección de la cinta y desplaza las señales de datos al siguiente conjunto de pistas. Esta operación continúa hasta que se hayan utilizado todas las pistas, momento en el cual la grabadora se apaga por sí misma.

Superposición. La superposición o secuenciación de grabadora a grabadora permite que dos grabadoras puedan ser operadas como una sola, continuamente y sin pérdida de datos. A medida que la grabadora en funcionamiento llega al "extremo de la cinta utilizable", la que se encuentra esperando inicia su funcionamiento a la velocidad adecuada cuando la primera grabadora se apaga. La cinta se rebobina en el aparato que ahora se encuentra esperando, lo cual la prepara para recibir la señal de secuenciación de la máquina en funcionamiento en cuanto ésta llegue al "extremo de la cinta utilizable". Por lo general, las señales de datos se conectan en paralelo a ambas máquinas.

Adaptador para cinta sin fin. La cinta sin fin se emplea para objetivos especiales de instrumentación. Puede utilizarse con cualquier tipo de accionamiento de eje de transporte y operar a tensión con un rodillo inversor de vuelta completa, almacenarse en un estuche o cassette sin arrollamiento ("al azar"), u operar en un ciclo circular de cinta lubricada. Algunos fabricantes ofrecen adaptadores para emplear cinta sin fin en aparatos ordinarios.

En la generalidad de las configuraciones suele disponerse de un almacenamiento de 5 a 150 pies. Estos adaptadores son ideales cuando los problemas de análisis de datos requieren la repetición continua de los datos en el equipo de análisis o para cualquier aplicación en la que se necesite la grabación y continua reproducción de datos sin necesidad de detener y rebobinar la cinta.

Señal de actuación. Como implica el nombre, una señal inicia la operación de grabación cuando excede de algún nivel umbral preestablecido. En caso de que la grabadora no sea capaz de adquirir la velocidad a tiempo para atrapar la señal también es posible utilizar un aparato de cinta sin fin. Con el ciclo de cinta operando constantemente en la sucesión grabación, reproducción y borrado, y el retraso de la cinta entre la cabeza de grabación y la cabeza de reproducción, la repentina aparición de una señal permite el retraso de cinta necesario para que la grabadora de datos adquiera la velocidad requerida. A continuación, la señal se transfiere a la grabadora de datos a partir de la cabeza de reproducción del aparato de cinta sin fin. Al salir de la cabeza de reproducción de este último aparato, la cinta se borra y la operación continúa.

45.2.5 Cinta magnética

Hace aproximadamente 30 años la cinta de grabación era una banda de papel recubierta con un material crudo de óxido rojo. Actualmente la cinta es una banda delgada de plástico, que suele ser de poliéster (Mylar), sobre la cual se coloca una emulsión de óxidos magnéticos altamente refinados. El aglomerante necesario para retener los óxidos también contiene una variedad de agentes químicos a fin de reducir la fricción, mejorar las características contra la humedad de las partículas de óxido, reducir la carga estática, lubricar la cinta, etc. El alma o base de la cinta se elabora en anchos rodillos continuos (en general, de 24 a 26 pulg), y después de que el aglomerante (que contiene el óxido) se ha colocado y curado sobre ella, se corta a la anchura deseada de la cinta con máquinas cortadoras de precisión.

Una cinta para instrumentación debe tener un grosor global bastante uniforme, así como una tolerancia muy restringida en la anchura. Por lo general, la tolerancia de anchura (también denominada tolerancia de corte) es $+ 0.000$ y $- 0.004$ pulg. El recubrimiento no debe rozar las cabezas o adherirse a otras partes del mecanismo de transporte de la cinta, incluso a altas velocidades de ésta. La distribución de las partículas de óxido debe ser constante en todo el grosor de la capa de óxido, así como en toda la longitud de la cinta. El tamaño de las partículas de óxido y las características magnéticas deben controlarse bastante bien en el caso de las cintas que serán utilizadas para la reproducción de longitudes de onda muy cortas.

Las características mecánicas de una cinta están, esencialmente, regidas por el material básico. Una base de cinta de poliéster presentará las siguientes propiedades:

1. Su absorción de humedad es la más baja de todos los materiales para la base y, por consiguiente, no presenta adhesión de capa con capa en climas húmedos.
2. Su resistencia a los hongos es excelente, mientras que la de los compuestos a base de celulosa es extremadamente baja.
3. Es posible utilizarla en un intervalo bastante amplio de temperatura, de -60 a 150 °C.
4. Su estabilidad dimensional la hace bastante superior a cualquier otro material básico.

La principal desventaja del poliéster es que se alarga en vez de romperse cuando se somete a esfuerzos superiores a su resistencia de fluencia. Si esto llega a ocurrir, toda la porción de la cinta y los datos comprendidos en ella deben descartarse. Sin embargo los mecanismos actuales de transporte de la cinta manejan ésta con bastante suavidad, y el alargamiento es bastante poco común, a menos que dicho mecanismo esté desajustado. Como resultado, las bases de poliéster se utilizan de manera casi universal en los aparatos para instrumentación, debido a que la diferencia de costo es bastante pequeña en términos del valor de los datos por grabar.

Tipos de cinta

La elección de la cinta para alguna aplicación específica ha representado un problema para los usuarios de las grabadoras de cinta para instrumentación. El rendimiento óptimo de cualquier aparato suele especificarse para un tipo particular de cinta, y el empleo de otra cinta puede afectar seriamente la capacidad del aparato para operar conforme a las especificaciones del fabricante. El documento IRIG 106-73 cubre la norma para cinta magnética de 1/2 y 1 pulg.

Los problemas de elección de cinta se han complicado con el advenimiento de las grabadoras de cinta para instrumentación de 1/4 pulg. A diferencia de las cintas de 1/2 y 1 pulg, que se elaboran principalmente para aplicaciones de instrumentación, la mayor parte de las cintas de 1/4 pulg se fabrican para ser utilizadas en equipo de audio. Dada la lista de todas las cintas existentes de 1/4 pulg, donde una cinta de alta calidad para instrumentación puede costar 10 veces lo que una cinta de inferior calidad para audio, elegir la cinta idónea para una aplicación específica puede volverse una decisión económica, así como de rendimiento. Para facilitar la referencia, seguidamente se presenta una combinación de las variables de la cinta en sí y la forma en que pueden afectar el rendimiento.

1. **Recubrimiento de óxido.** Las diferencias en las características del óxido y en el grosor del recubrimiento pueden provocar un pico hasta de 8 dB o un descenso en la respuesta de frecuencia de una grabadora. El mejor método para minimizar este problema es utilizar los tipos de cinta para los que fue diseñada la grabadora.
2. **Respuesta de frecuencia.** Para grabadoras de ancho de banda intermedio (hasta 0.25 mpulg de longitud de onda) es posible utilizar cintas de calidad para audio con resultados satisfactorios. Sin embargo, la variación en las características de respuesta de frecuencia provocarán que varíe bastante el rendimiento de diferentes tipos de cinta. Las cintas de audio que funcionan bien de manera consistente comparadas con las cintas para instrumentación son tan costosas como éstas; tienen la ventaja de conseguirse con facilidad en tiendas locales al menudeo. Una desventaja importante de las cintas para audio es que los parámetros importantes de la cinta no se especifican o no están garantizados por el fabricante. En esta industria tan dinámica, la existencia continua de una cinta de tipo específico para audio es una suposición incierta.
3. **Efecto en el desgaste de la cabeza.** En algunas de las cintas para audio más económicas, no se emplea un método adecuado para adherir el aglomerante del recubrimiento de óxido sobre el material base de la cinta. El óxido que se desprende como resultado de lo anterior provoca desgaste excesivo tanto en las cabezas como en las guías de la cinta. A la vista del costo de reemplazar cabezas de grabadora, el gasto en cinta de mayor calidad bien puede reducir el costo de vida útil de un mecanismo bastante utilizado.

Tabla 45-4. Características de la cinta

Parámetros de tipo	Audio		Instrumentación		
	Bajo costo	Alto costo	Banda intermedia	Banda ancha	Dióxido de cromo
Características de desgaste de la cabeza	Altamente abrasiva	Bajas	Bajas	Muy bajas	Altas
Mínima longitud de onda (λ) grabada ^a	> 0.25 mpulg sólo	> 0.25 mpulg sólo	> 0.25 mpulg sólo	> 0.0625 mpulg	> 0.0625 mpulg
Señal-ruido @ $\lambda > 0.25$ mil	Deficiente	Aceptable	Aceptable	De deficiente a aceptable	Excelente
Señal-ruido @ $\lambda < 0.25$ mil	Deficiente	Deficiente	Deficiente	Aceptable	Aceptable
Confiabilidad de funcionamiento de la cinta	Deficiente	Deficiente	Aceptable	Aceptable	Aceptable

^a Frecuencia = velocidad de la cinta/ λ .

En la generalidad de las cintas se emplea una partícula de óxido ferroso, mientras que, cuando el objetivo principal es una mayor respuesta de frecuencia se utiliza un dióxido de cromo más duro y permeable. El uso de esta cinta de dióxido de cromo en aparatos con cabezas de mumetal en vez de cabezas de ferrita provocará un rápido incremento en el ritmo de desgaste de la cabeza.

En la tabla 45-4 se enumeran algunos tipos de cintas, sus características y la forma de comportamiento comparadas entre sí.

Aunque las cintas de calidad superior se venden a un mayor precio, no representan una respuesta total a todo problema de medición. Existen en el mercado grabadoras para instrumentación en las que se especifica cinta para audio a fin de satisfacer las especificaciones publicadas, y deben considerarse en cualquier análisis que se emprenda.

Caídas de señal

Un problema universal en la grabación directa para instrumentación son las caídas de señal o fallos momentáneos de registro. Una caída se define como una reducción del 50 % (o más) en la amplitud de los datos reproducidos. Aunque por lo general son provocadas por contacto deficiente entre la cabeza y la cinta durante el proceso de grabación o de reproducción, también pueden ser provocadas por una distribución deficiente de partículas del óxido en el aglomerante. Los efectos de un contacto deficiente entre cabeza y cinta pueden reducirse limpiando las cabezas, guías y piezas de la grabadora que entran en contacto con la cinta, pero no hay nada que pueda hacer el usuario con respecto a una cinta que tenga una mala distribución de óxido o una superficie áspera.

La atenuación de la señal que resulta de un contacto deficiente entre cinta y cabeza (o de una superficie áspera de la cinta) es más pronunciada en las longitudes de onda más cortas, mientras que la ausencia de material magnético en el aglomerante afecta todas las

frecuencias. El efecto del contacto deficiente entre cabeza y cinta se denomina pérdida por separación, y puede ilustrarse por medio de:

$$\text{Caída en el nivel de reproducción (dB)} = 54 \frac{d}{\lambda}$$

donde d es la separación de la cinta con respecto a la cabeza y λ es la longitud de onda grabada, ambas en pulgadas.

En la figura 45-25 se muestra la atenuación como función de la razón separación-longitud de onda. Ésta es una curva universal, aplicable a cualquier velocidad, frecuencia y separación. Debe observarse que esta figura ilustra la pérdida de nivel de la señal sólo en la reproducción, suponiendo que el contacto entre la cabeza y la cinta es adecuado durante la grabación. Es obvio que en la grabación ocurrirá una pérdida de señal parecida si existe una separación entre la cinta y la cabeza correspondiente. Además de tales pérdidas dependientes de la frecuencia, las frecuencias muy polarizadas son perturbadas por la pérdida por separación, hasta provocar que todas las componentes de frecuencia de la señal de datos sufran atenuación o deformación.

En la reproducción de una señal de baja frecuencia con longitud de onda de 15 mpulg, una separación de 1 o 2 mpulg apenas afecta el nivel, pero a una alta frecuencia de 1 mpulg de longitud de onda, incluso una separación de 0.5 mpulg dará por resultado una caída de más de 20 dB. (Una longitud de onda de 1 mpulg resulta cuando se graban 15 000 Hz a 15 pulg/s [ips], 60 000 Hz a 60 pulg/s [ips], etc.) En grabación para instrumentación, una caída de 6 dB o del 50 % en la señal se considera crítica. Con una longitud de onda de 1 mpulg, esto ocurre con 111 μ pulg (0.111 mpulg de separación entre la cabeza y la cinta). Dado que una partícula de polvo tiene aproximadamente este tamaño, la posibilidad de pérdidas por separación y la importancia de mantener extremadamente limpia la

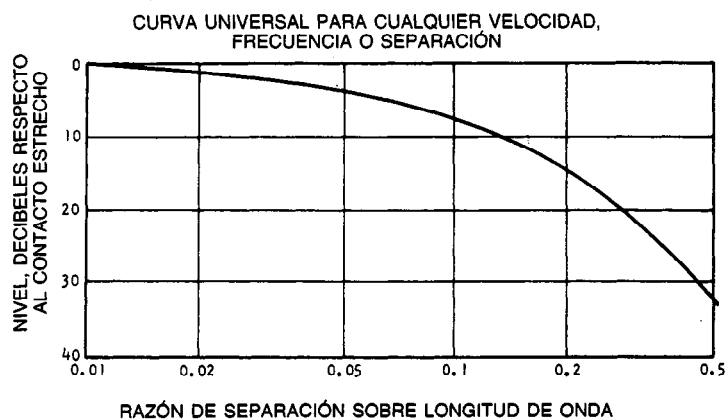


Fig. 45-25. Atenuación de la señal provocada por contacto deficiente entre la cinta y la cabeza durante la reproducción.

grabadora son evidentes. Cuando las cabezas dejan de limpiarse durante mucho tiempo es fácil que ocurra la pérdida total de las altas frecuencias, y también es posible culpar a cualquier parte de la grabadora.

Manejo de la cinta magnética

Cuando se maneja la cinta, como durante el empalme, las manos del operario deben estar completamente limpias a fin de evitar contaminar la cinta con grasas y sales del cuerpo, que podrían provocar la adhesión de partículas extrañas. Especialmente, no se recomienda utilizar *masking tape* o cinta adhesiva ("celofán scotch") como empalmador o para sujetar el extremo de la cinta, ya que pequeños depósitos del adhesivo se pegan a la cinta magnética.

Es necesario limpiar las cabezas y guías, a fin de eliminar acumulaciones de materiales extraños, cada vez que se coloque una cinta en la grabadora. Debe seguirse al pie de la letra el procedimiento de limpieza del aparato sugerido por el fabricante. Si los datos por grabar son de extrema importancia, debe recordarse que una cinta que se ha hecho pasar varias veces por el mecanismo de transporte tendrá menos caídas de señal que una nueva, debido a la acción resultante de pulido o alisado.

Limpieza

Si existen caídas de señal debido a la contaminación por polvo, es necesario limpiar con cuidado la superficie y parte posterior de la cinta con un paño libre de pelusa, como ante muy suave.

A fin de deshacerse de contaminación que no se elimina con facilidad debe emplearse un paño ligeramente humedecido con freón TF. También es posible utilizar solventes del tipo de los hidrocarburos alifáticos (heptano, gasolina, nafta), pero se ha de tener cuidado, ya que son inflamables. El freón TF no es tóxico ni inflamable. No debe utilizarse tetracloruro de carbono, alcohol etílico, tricloroetileno o cualquier otro agente limpiador desconocido, ya que pueden ablandar el óxido, deformar la parte de soporte de la cinta o ambas cosas.

Almacenamiento

Cuando no se encuentran en uso, las cintas deben colocarse en un carrete de precisión, arrolladas de manera uniforme a tensión moderada, y almacenarse a fin de protegerlas. La tensión de rebobinado recomendada para la mayor parte de las grabadoras en las que se utiliza cinta de 1 mpulg es de 4 a 5 oz por cada 0.25 pulg de anchura de la cinta. El mejor método de almacenamiento es colocar el carrete de la cinta en una caja hermética de plástico y guardarla verticalmente en un anaquel con divisiones para cada carrete. La caja de plástico protege la cinta de polvo y cambios repentinos de humedad y temperatura, y también protege la cinta y el carrete de daño durante su manipulación cuando se lleva del área de almacenamiento al área de trabajo o viceversa. La bolsa de plástico y la caja de cartón en que se embala la cinta probablemente serán de mayor utilidad que contenedores especiales, y representan una protección razonable para muchas condiciones.

Si es necesario almacenar las cintas en presencia de campos magnéticos, de CA o CD, existen contenedores especiales. Estos recipientes impiden que los datos se borren en cualquier campo magnético excepto en los extremadamente intensos. Sin embargo, es más aconsejable almacenarlas lejos del campo.

Deben evitarse los extremos de temperatura y humedad. Por lo general, las condiciones recomendadas para el almacenamiento de ambos tipos de bases son:

Humedad relativa	40 a 60 %
Temperatura	60 a 80 °F (16 a 27 °C)

Si durante el almacenamiento o transporte se encuentran con temperaturas extremas, es necesario hacer que la cinta se encuentre en equilibrio antes de utilizarla. Suponiendo, por ejemplo, que una cinta ha estado almacenada o se ha transportado a temperaturas bajo cero, debe almacenarse por lo menos de 4 a 8 horas a temperatura ambiente antes de poder emplearla, aunque en realidad no se encontrará en equilibrio completo sino hasta después de aproximadamente 16

horas. Este tiempo puede acortarse elevando la temperatura ambiente, pero la temperatura no debe exceder los 100 °F (38 °C); en caso contrario se formará condensación sobre la cinta, lo que también es un problema. Debe evitarse el empleo de calor directo, como el de lámparas incandescentes u otros calentadores para elevar la temperatura de la cinta.

Las condiciones de temperatura y humedad durante el embarque de la cinta son impredecibles, especialmente cuando el envío es aéreo. La temperatura y humedad varían de forma tan rápida durante el despegue que los esfuerzos de un lado a otro o a través del rollo de cinta no llegan a equilibrarse antes de que estas condiciones se inviertan durante el aterrizaje. Un alargamiento no uniforme en el sentido transversal de la cinta da por resultado arrugas o rizos en los bordes al reproducirla. Los contenedores herméticos y térmicamente aislados de plástico reducen los efectos de temperatura y humedad y suelen permitir que la cinta alcance el equilibrio de temperatura con el método descrito en el párrafo anterior. En caso de requerirse máxima seguridad para la cinta, pueden emplearse cajas de cartón herméticas para el embalaje.

Si la cinta se ha sometido a combinaciones de temperatura y humedad que provocan adherencia entre sus capas, es posible que ya no pueda recuperarse. No debe colocarse la cinta en el modo de avance rápido cuando se quiera desbobinar un carrete sospechoso de almacenamiento o condiciones de transporte anormales; en vez de ello es necesario utilizar un modo de accionamiento a baja velocidad. A fin de obtener funcionamiento óptimo de la cinta es necesario desbobinar y arrollar todas las cintas que han estado almacenadas durante mucho tiempo.

Deformación física

Aunque la mayor parte de las caídas de señal en las grabadoras para instrumentación se deben a partículas de polvo y otros contaminantes que separan la cinta de la cabeza, otras dos causas importantes son indentaciones y pliegues en el material base. Las indentaciones pueden ser provocadas por partículas extrañas que se adhieren fuertemente en el rodillo o por asperezas en la superficie del cubo sobre el cual se arrolla la cinta, y pueden ser la causa de alguna abolladura o pliegue permanente en muchas capas de la cinta, que no es posible aplanar y alisar a medida que la cinta pasa por la cabeza. Los esfuerzos en el rodillo que bastan para alargar un 4 % la base de la cinta suelen dejar una impresión permanente; los esfuerzos por debajo del punto del 5 % no suelen ser permanentes. Los pliegues suelen ser provocados por el manejo de la cinta (arrollamiento, empalmes, retirar la cinta de las guías, etc.) o por daño en sus bordes debido a devanado no uniforme.

La mayor parte de las causas de deformación del material base pueden eliminarse mediante el empleo de un carrete de precisión. Un carrete de precisión típico tiene bridas rectas o ahusadas labradas y espaciadas con exactitud a fin de minimizar la dispersión de las vueltas durante el arrollamiento. El diseño de la brida también proporciona mucha mayor protección

contra el polvo y el aplastamiento de los bordes de la cinta. El cubo no tiene ranuras de bobinado que provoquen deformación de las vueltas interiores. En vez de ello, con frecuencia se cubre con una banda de fricción de neopreno a fin de ayudar al devanado. Este anillo actúa como cojín para las capas más interiores de la cinta y tiende a minimizar la deformación debida a la presión del arrollamiento y a los esfuerzos de dilatación-contracción.

Borrado o saturación

Las propiedades magnéticas de las cintas para instrumentación son indefinidamente estables. La retentividad magnética es permanente, a menos que se modifique por medios magnéticos. Por ejemplo, puede modificarse por medio de campos magnéticos de imanes permanentes o electroimanes. Es muy probable que dichos campos provoquen un borrado parcial si se colocan a pocos centímetros de la cinta.

Este principio se utiliza en el proceso de borrado masivo, en el cual se desmagnetiza todo un carrete de cinta sin necesidad de desbobinarlo. Sin embargo, los campos necesarios para producir el borrado total son tan intensos que no es probable que campos magnéticos parásitos provoquen problemas de este tipo. El borrado total (considerado para fines de este análisis como la reducción de la señal hasta un punto por debajo del nivel de ruido del sistema) no suele ocurrir a menos que el campo sea lo suficientemente intenso para ejercer en la cinta una atracción observable. Sin embargo, puede ocurrir ligero borrado sin atracción o vibración observables.

En la figura 45-26 se ilustra la relación entre la intensidad de campo y el borrado, según se observa en experimentos realizados con un borrador masivo de CA típico. Es observable algo de borrado a una intensidad de campo de sólo 100 oersteds. Por lo general una pérdida de 6 dB se considera crítica, debido a que representa un 50 % de reducción en la intensidad de la señal. En algunas aplicaciones una pérdida de 1 dB podría ser grave.

Tanto las cintas sin grabar como las grabadas deben mantenerse alejadas de borradores masivos electro-magnéticos y gabinetes de almacenamiento con pestillos magnéticos. Las cintas sin grabar no deben colocarse en la cercanía de campos magnéticos de CD, como tubos de ondas progresivas o imanes magnetron, debido a que pueden polarizarse fuertemente o incluso crear una grave deformación en el proceso de grabación (es decir, se reducirá la relación señal-ruido resultante).

Si algunas piezas de la grabadora se magnetizan, pueden provocar el borrado en la cinta, posible saturación de la cinta y degradación de la señal. Como medida preventiva se recomienda la desmagnetización periódica de las piezas críticas, especialmente las cabezas.

Para proteger una cinta grabada contra el borrado accidental durante el embarque, la cinta puede embalsarse con una separación considerable (tal vez con madera) entre ella y el cartón del embalaje. Esta separación es efectiva para reducir la posibilidad de bo-

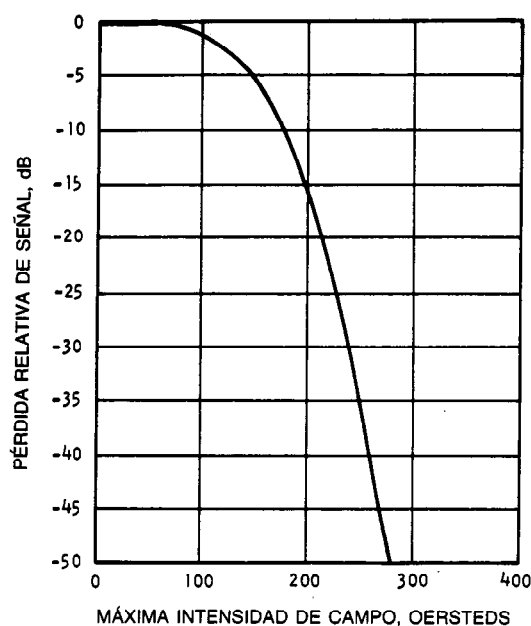


Fig. 45-26. Borrado como función de la intensidad de campo.

rrado accidental por campos encontrados durante el transporte, ya que la intensidad de campo varía inversamente con el cuadrado de la distancia. Suponiendo que durante el embarque no se encuentre una intensidad de campo mayor de 1 000 oersteds (esta hipótesis no es verificable, aunque sea razonable), una separación de 3 pulg entre la cinta y el cartón del embalaje proporcionará la protección adecuada.

El contenedor especialmente protegido que se describió en el párrafo sobre el almacenamiento puede emplearse en el embarque si durante el trayecto se espera encontrar grandes campos magnéticos parásitos. Sin embargo, la experiencia indica que las cajas normales de embalaje suelen ser satisfactorias.

LIBROS Y REVISTAS

- S. Bannerjee, *Audio Cassettes: The User Medium*, UNESCO.
- B. B. Bauer, F. A. Comerci, E. J. Foster y A. J. Rosenbeck, "Audibility of Tape Dropouts", *JAES* 15:147-151.
- R. M. Dolby, *Audio Noise Reduction System*, Dolby Labs., Londres.
- A. Everest, *Handbook of Multichannel Recording*, Tab Books, Blue Ridge Summit, PA.
- F. Jorgenson, *The Complete Handbook of Magnetic Recording*, Tab Books, Blue Ridge Summit, PA.
- I. Levine y D. E. Daniel, "Determination of the Recording Performance of a Tape from Its Magnetic Properties", *JAES* 7:181-188.
- J. G. McKnight, "Erasure of Magnetic Tape", *JAES* 11:223-233.
- J. G. McKnight, "Mechanical Damping in Tape Transports", *JAES* 12:140-146.
- J. G. McKnight, "Tape Reproducer Response Measurements with a Reproducer Test Tape", *JAES* 15:152-156.
- A. H. Morris, "A New Magnetic Tape with Greater Dynamic Range", *JAES* 12:36-39.
- R. E. Runstein, *Modern Recording Techniques*, Howard W. Sams, Indianapolis, IN.
- W. G. Salm, *Cassette Tape Recorders, How They Work*, Tab Books, Blue Ridge Summit, PA.
- A. A. Sarati, "A High Resolution Stereo Magnetic Head for Four Track Application", *JAES* 8:243-245.

45.3 ALTAVOCES

Albert B. Grundy

45.3.1 Introducción

Para la comprensión de los altavoces se requieren por lo menos tres campos de conocimiento independientes:

1. Los transductores, es decir, los principios en que se apoyan los elementos que generan el sonido.
2. Los muebles acústicos, o sea, la forma en que éstos y las pantallas acústicas (bafles) afectan el sonido del sistema.
3. La acústica de recintos; esto es, la influencia de la sala o salón sobre la respuesta del altavoz. Es necesario saber cómo medir la respuesta de un altavoz, tanto anecoicamente como en una sala real, cómo interpretar las mediciones y cómo corregir de forma apropiada las deficiencias de la respuesta.

45.3.2 Tipos de transductores

Altavoces dinámicos

Altavoz de cono. La mayor parte de los transductores son motores electrodinámicos, de los cuales el más común suele denominarse altavoz PM, que se ha vuelto casi una denominación genérica. Aquí se prefiere el nombre "altavoz de cono" a fin de diferenciarlo de su pariente más cercano, el "motor electroacústico de compresión".

El altavoz de cono se construye casi de la misma forma que el micrófono de bobina móvil, ya que ambos dependen de bobinas móviles dentro de un campo magnético. Con respecto a la figura 45-27, un altavoz de cono puede ser cualquiera cuyo diámetro mida de 1 hasta 30 pulg. El cono (o diafragma) se encuentra adherido a un material exterior flexible, denominado borde o suspensión periférica, que es elástico pero

también ayuda actuando como resorte de centrado. En el centro, el diafragma se une a la bobina móvil. La "araña" (centrador flotante del cono) mantiene centrada la bobina móvil en el entrehierro magnético y evita que roce en los lados. También proporciona la necesaria fuerza elástica de restitución que hace regresar el cono a su posición de reposo.

Aunque la bobina móvil puede desplazarse con libertad hacia atrás y adelante, su recorrido debe permanecer dentro de las líneas de flujo del imán, ya que en caso contrario puede producirse una deformación no lineal. Un método para garantizar la linealidad es hacer la bobina móvil más larga que el entrehierro, de modo que la misma cantidad de cobre siempre se encuentre inmersa en el campo dentro de un intervalo dado del desplazamiento. Con este método se sacrifica algo de eficiencia.

Algunos altavoces de cono (denominados coaxiales) tienen más de un diafragma alrededor de un solo eje. Aunque el cono circular es el más común, en altavoces de cono de alta calidad también se han utilizado formas rectangulares o irregulares. Los conos solían elaborarse de papel o pulpa de celulosa, pero en la actualidad la mayor parte se fabrican en materiales sintéticos, como polipropileno o bextreno, o bien de metales ligeros, por lo general aluminio. Dos ventajas de los materiales opcionales sobre el papel son una menor susceptibilidad a la humedad (el cono de papel higroscópico es más lento en ambientes húmedos) y mayor rigidez sin aumento de peso.

En aplicaciones comerciales, los altavoces de cono se utilizan para reproducción tanto de baja como de alta frecuencia, pero por lo general no se ha demostrado que sean lo suficientemente robustos o eficientes para aplicaciones profesionales de alta frecuencia. En condiciones difíciles (p. ej., altos niveles de presión del sonido), suele emplearse una combinación de motor de compresión y bocina, en vez de un altavoz de cono (de radiación directa). Sin embargo, con la demanda creciente para alta fidelidad y menos distorsión en la sala de control, es posible encontrar altavoces más resistentes sustituyendo al tradicional motor de compresión. A fin de obtener más salida, también es posible acoplar un altavoz de cono a una bocina.

Motor de compresión. Como transductor, el motor electroacústico de compresión es primo del tipo cónico. La principal diferencia es que el cono se sustituye por un diafragma más pequeño, usualmente de aluminio. Un motor de compresión siempre se adapta a la garganta de una bocina especial, que acopla la impedancia acústica del motor con el aire (fig. 45-28).

Altavoces magneplanar y de cinta. De uso exclusivo para el hogar, el altavoz magneplanar es un diseño único en el que se utiliza un diafragma de poca masa de poliéster, como el del altavoz electrostático, pero se produce un campo magnético estático por muchos imanes delgados en forma de barras. Los alambres que actúan como bobina móvil se encuentran a través del delgado panel de poliéster. El grosor total del radiador bipolar es de 2 pulg.

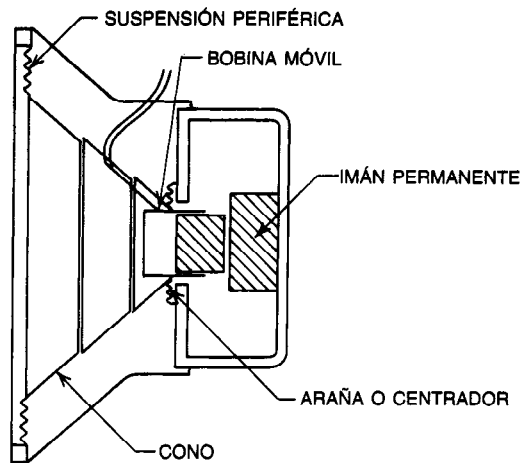


Fig. 45-27. Altavoz de cono.

El altavoz de cinta es parecido al micrófono de cinta. Produce excelente respuesta de alta frecuencia, pero a niveles limitados de presión acústica. Aunque la tecnología actual permite el uso de altavoces de cinta para altas frecuencias (*tweeters*) y para la mitad superior del intervalo, nadie ha producido una cinta práctica capaz de reproducir bajas frecuencias.

Altavoz electrostático

Este altavoz es el análogo del micrófono de condensador (capacitor). Dos placas estacionarias operan en un mecanismo de acción alternada a fin de minimizar la distorsión de segundo armónico. Los dos electrodos constan de una rejilla fina para dejar pasar el sonido (fig. 45-29). Dado que la polaridad de los electrodos es opuesta, el diafragma delgado inerte (que suele ser de poliéster claro) es atraído de forma alternada por el polo electrostático frontal y por el posterior. Para impedir que el diafragma atravesase el dieléctrico o el entrehierro de aire y se "pegue" a cualquiera de los polos, una serie de espaciadores delgados mantienen centrado el diafragma entre los electrodos.

El intervalo dinámico del altavoz electrostático está determinado por el voltaje de polarización, que debe limitarse de 1 000 a 2 000 V aproximadamente a fin de evitar la formación de arcos o chispas. Su empleo está limitado al hogar debido a su intervalo dinámico relativamente reducido y su patrón de radiación bipolar.

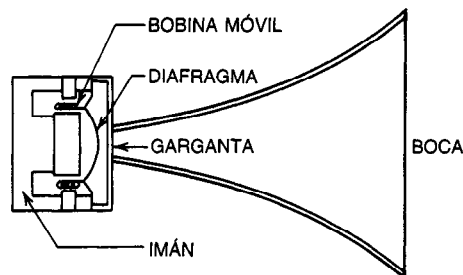


Fig. 45-28. Motor electroacústico de compresión y bocina.

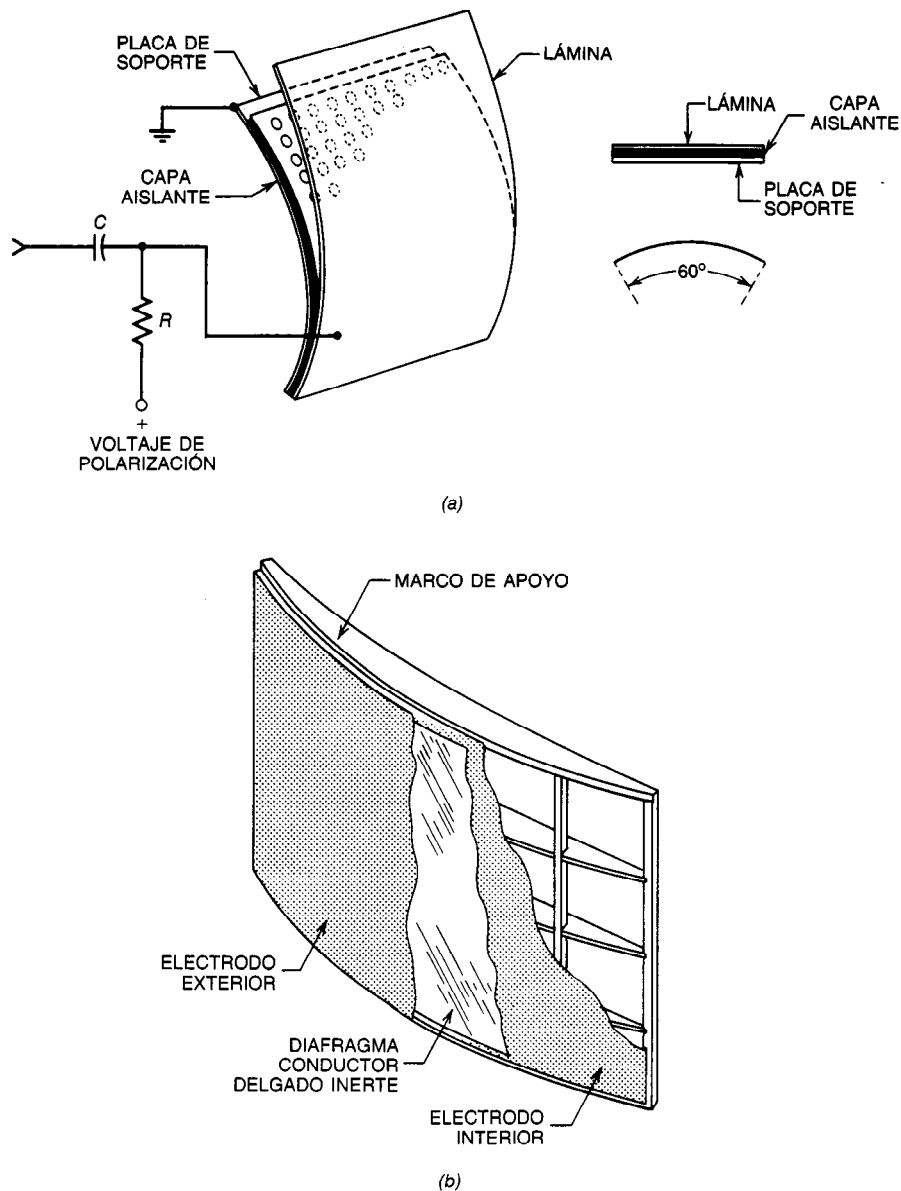


Fig. 45-29. a) Altavoz electrostático. b) Vista de un altavoz electrostático en la que se observa la construcción interna. Cortesía de Pickering and Co. Inc.

Transductor piezoeléctrico

Análogo al micrófono de cerámica, en el altavoz piezoeléctrico se emplea un elemento de transducción que suele constar de materiales cerámicos policristalinos. El diafragma de un altavoz piezoeléctrico está conectado al transductor por medio de varillaje (reenvío) mecánico (fig. 45-30). Posteriormente, suele acoplarse con una bocina al aire.

Las capacidades de desplazamiento mínimas de los transductores piezoeléctricos han restringido su empleo a aplicaciones de alta frecuencia y bajo nivel. Sin embargo, muchas empresas de refuerzo de sonido para

concreto (hormigón) vibrado han producido columnas de sonido prácticas con múltiples transductores piezoeléctricos.

Su aplicación profesional ha sido posible gracias a su bajo precio y elevada impedancia interna (lo que permite conexiones en paralelo).

Otros tipos

Ya se han construido transductores experimentales que algunas veces se han colocado en el mercado para el consumidor. Algunos de ellos son el altavoz esotérico de "llama" y el altavoz iónico.

45.3.3 Tipos de muebles acústicos

La mayor parte de los altavoces son del tipo híbrido, en los que se emplean combinaciones o variaciones de los muebles acústicos que se describen a continuación.

Pantalla acústica (baffle)

Un altavoz en un entorno al aire libre debe contar necesariamente con un corte de baja frecuencia (fig. 45-31), determinado por las longitudes de onda de la señal y la diferencia en la longitud de la trayectoria entre las partes posterior y anterior del altavoz excitador. La adición del baffle agranda la diferencia en la trayectoria entre las partes anterior y posterior y reduce la frecuencia de cancelación. Por supuesto, el último baffle es el baffle infinito, como, por ejemplo, un altavoz montado en una pared entre dos grandes salas.

Carga por bocina

Incluso antes de que Rice y Kellogg inventaran el altavoz dinámico en 1925, varios tipos de bocinas proporcionaban amplificación acústica en receptores telefónicos. En todos los sistemas profesionales diseñados para producir altos niveles de presión acústica con elevada eficiencia es posible encontrar variaciones modernas del altavoz de bocina. Con el mayor conocimiento sobre los principios físicos implicados, se han reducido considerablemente la deformación y la calidad resonante globales de los altavoces de bocina. De hecho, los ingenieros están observando una verdadera convergencia en las calidades de grandes sistemas tipo profesional de carga por bocinas y de los pequeños altavoces tipo casero de radiador directo. Por ejemplo, mientras están aumentando la eficiencia y el manejo de potencia de las unidades caseras, también están mejorando la respuesta de frecuencia y disminuyendo la distorsión de las unidades profesionales.

En la figura 45-28 se muestra un bocina exponencial. Una razón de abertura (abocinado) exponencial proporciona un rápido incremento en la resistencia de la garganta a frecuencias inferiores para una longitud de bocina dada. Otros tipos de razones de abertura son la parabólica, cónica e Hypex. Los factores que determinan la longitud de una bocina exponencial son la razón de abertura, la pretendida baja frecuencia de corte, la velocidad del sonido en el aire y el área de la garganta de la bocina en el punto en que se encuentran el altavoz excitador y la bocina.

La resonancia de la bocina, que provoca aberraciones y distorsión en la respuesta de frecuencia, es una de las consideraciones fundamentales del ingeniero. Si la campana de la bocina se elabora grande respecto a la frecuencia más baja por reproducir, se reducen la resonancia y la distorsión. También es responsabilidad del diseñador colocar un filtro eléctrico separador de frecuencias antes de la unidad a fin de limitar las frecuencias de entrada a los límites de diseño del sistema acústico de bocina y altavoz excitador.

Las bocinas reentrantes o plegadas siguen el principio de muchos instrumentos musicales de metal, es decir, cuentan con mayor área total y menor frecuencia de corte, pero menor longitud física (fig. 45-32).

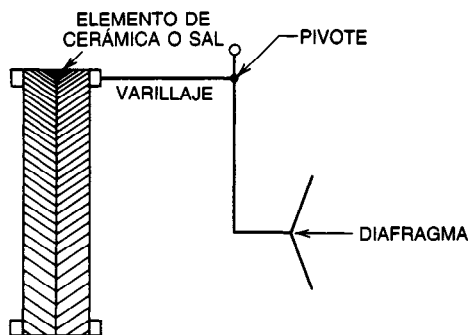


Fig. 45-30. Mecanismo de excitación de un altavoz de cerámica o de cristal. Cortesía de Audio Cyclopedia.

Caja acústica con escape de bajos

Otras denominaciones de la caja acústica (baffle) con escape de bajos son las de reflector de bajos e inversor de fase acústica. Introducido por primera vez hacia 1937, este dispositivo fue diseñado para incrementar la respuesta de baja frecuencia sin menoscabo de la eficiencia, conforme a los principios del resonador de Helmholtz. Aproximadamente antes de 1971 el diseño de tales sistemas era más bien un "arte de magia" que una ciencia. Sin embargo, en dicho año, A. N. Thiele publicó una serie de documentos que marcaron un hito en el diseño de la caja acústica con escape de bajos^{1,2}. Su trabajo fue mejorado por R. H. Small en 1972 y 1973^{3,4}. Antes de Thiele y Small se pensaba que la caja sellada era inherentemente más lineal que la caja con escape de bajos, aunque ahora se sabe que ambos sistemas pueden elaborarse con respuesta de frecuencia lineal. Así, los diseñadores actuales de bocinas eligen los principios de diseñar con base en factores empíricos, incluyendo eficiencia, tamaño de abertura, volumen del mueble, resonancia del altavoz, frecuencia inferior de corte y distorsión armónica. Actualmente la mayor parte de estos factores pueden modelarse en una computadora mucho antes de cortar la madera para la caja. Gracias a Thiele y Small, un sistema de caja acústica con escape de bajos diseñado de manera apropiada ya no merece el apodo de "caja retumbante".

El principio de la abertura de escape de resonancia es actuar como resonador, extendiendo la respuesta de baja frecuencia del altavoz por debajo de la frecuencia de resonancia del cono (fig. 45-33). Por lo general la abertura es un simple orificio de dimensiones calculadas en el frente o en un lado del mueble. Es interesante observar que el altavoz con abertura puede considerarse incluso una subcategoría del altavoz cargado por bocina, ya que si se empieza con un diseño con abertura y se construye un laberinto acústico dentro del mueble cuya sección transversal aumente de forma gradual y termine en la "boca" de la abertura, se ha construido una bocina plegada de baja frecuencia.

Caja sellada

El mueble acústico de suspensión neumática, que se dice fue diseñado por Edgar Vilchur a mediados de la década de 1950, es conocido por su profunda respuesta

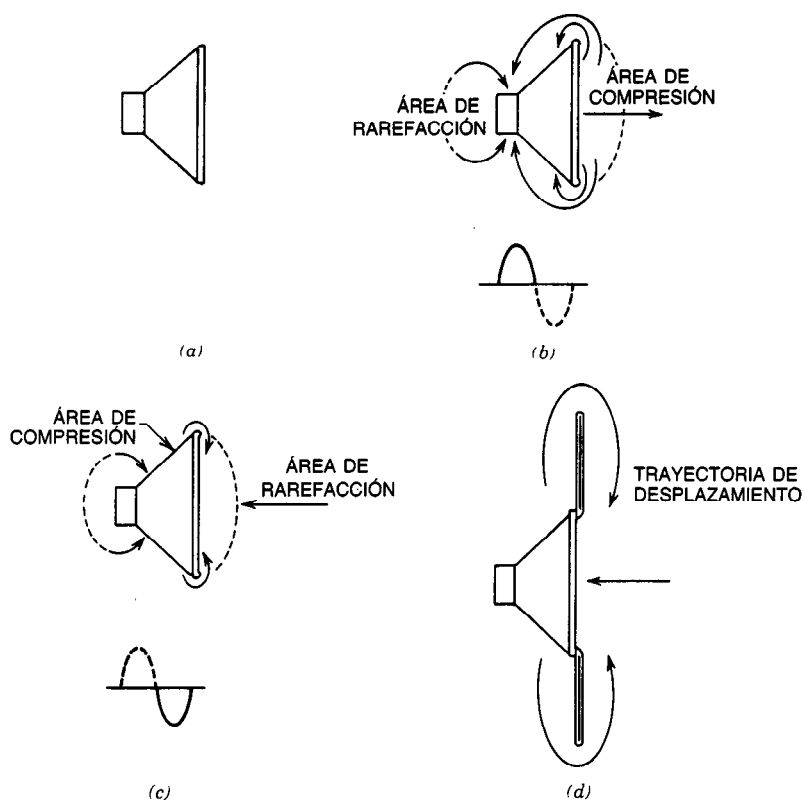


Fig. 45-31. Altavoz que opera sin baffle; se observa cómo se cancelan las ondas sonoras frontales y posteriores a las frecuencias inferiores: a) diafragma en reposo; b) el diafragma se mueve hacia delante; c) el diafragma se mueve hacia atrás; d) bocina montada en un baffle; se observa el aumento en la longitud del desplazamiento provocado por el baffle. Cortesía de Audio Cyclopedia.

bien amortiguada de bajos reducidos hasta un límite definido por el volumen interior del mueble acústico. Es un dispositivo de eficiencia relativamente baja que sólo se popularizó cuando aparecieron en el mercado los amplificadores de alta potencia. En un altavoz de suspensión neumática se emplea la combinación de un mueble amortiguado sellado, de volumen específico, con un altavoz que posee una suspensión periférica muy flexible.

Este altavoz más dócil se diseña para una frecuencia de resonancia casi subsónica al aire libre, por lo cual es casi completamente no amortiguada. Aun así, la adición de un mueble sellado y amortiguado (con lana o fibra de vidrio) actúa como resorte, de hecho un resorte más lineal que el que se obtendría con un altavoz poco dócil sin ayuda. Se ha afirmado que el resorte lineal proporcionado por el aire dentro del mueble sellado produce menor distorsión armónica de graves que un sistema de caja con escape de graves, pero avances recientes han hecho que los dos sistemas sean por completo competitivos. El altavoz de suspensión neumática tiene poco o ningún aumento de respuesta de frecuencia o impedancia y una caída rápida por debajo del punto de corte de baja frecuencia. La eficiencia se ha medido aproximadamente en el 0.5 al 2 %.

45.3.4 Filtros separadores de frecuencias

El típico altavoz excitador de cono de diámetro apreciablemente grande es capaz de reproducir altas frecuencias, pero no tan bien como las bajas. Cuando un gran altavoz excitador reproduce las altas frecuencias, sólo vibra una parte del diafragma, y se pone en peligro la correcta difusión de la alta frecuencia. Por consiguiente, al altavoz de bajas frecuencias (*woofer*) no suelen alimentarse las altas frecuencias. De manera parecida, el altavoz para las altas frecuencias (*tweeter*) puede ser dañado por las bajas frecuencias, que podrían llevarlo más allá de sus límites de desplazamiento, y se evita que lleguen a él bajas frecuencias. La red (circuito) divisora o separadora se denomina *filtro separador*.

Los filtros separadores pueden ser activos o pasivos. El término "filtro separador pasivo" suele referirse a ciertos componentes reactivos y resistivos localizados entre el amplificador de potencia y los altavoces. El término "filtro separador activo" suele referirse a componentes reactivos y resistivos localizados antes de la entrada de un amplificador de potencia o (rara vez) dentro del ciclo de retroalimentación negativo del amplificador de potencia. Cuando se utiliza un filtro se-

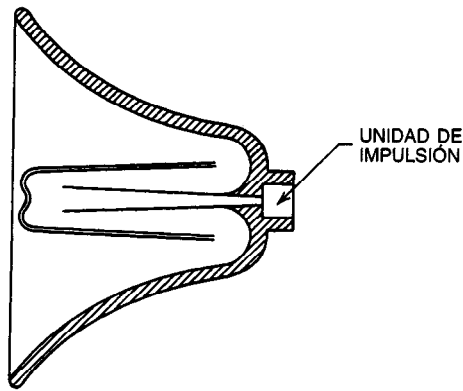


Fig. 45-32. Bocina plegada.

parador activo, un amplificador de potencia separado alimenta cada altavoz o grupo de altavoces en el sistema de altavoces múltiples.

Deben calcularse con bastante cuidado las razones de las pendientes y respuestas de frecuencia de los filtros separadores a fin de producir la mejor respuesta global de todo el sistema y garantizar que los altavoces no sean sobreexcitados. Los diseñadores han empleado pendientes de 6 dB a 18 dB por octava, y algunas veces más.

45.3.5 Patrones polares

En la figura 45-34 se muestran los patrones polares de un altavoz típico. Cuando se emplean altavoces múltiples, estos patrones polares pueden ser más bien com-

plicados. Los diseñadores deben tomar precauciones a fin de obtener una respuesta de frecuencia uniforme dentro de un área prescrita de escucha horizontal y vertical.

45.3.6 Técnicas de medición

Los primeros diseñadores y algunos de los actuales ingenieros a menudo miden la respuesta de frecuencia de un altavoz en el exterior, a fin de eliminar la influencia negativa de las reflexiones que hay en una sala de audición. Muchos incluso entierran el mueble de la bocina hasta que su cara se encuentra al ras del piso, eliminando de esta forma la mayor parte de los efectos de difracción audibles y medibles. La cámara anecoica está diseñada para el mismo fin, pero la generalidad de las cámaras de tamaño permisible o práctico no son lineales a bajas frecuencias debido a ondas estacionarias.

Las curvas de respuesta registradas dentro de una cámara anecoica raramente parecen correlacionarse con las que escucha el oído, incluso dentro de una "excelente" sala de audición. Muchos altavoces que presentan respuesta lineal de alta frecuencia dentro de la cámara anecoica suenan muy brillante o agudo cuando se operan en una sala normal. Al menos una de las causas de esto es que la combinación oído-cerebro integra en alguna medida el sonido directo de un altavoz con los sonidos reflejados por las paredes y el piso. Si, por ejemplo, la respuesta de alta frecuencia fuera del eje del altavoz no es paralela a su respuesta anecoica, atenuándose mucho más rápido, probablemente el altavoz tendrá un sonido más brillante para el oyente.

Esto ilustra sólo uno de los muchos problemas asociados con la medición de las respuestas de un al-

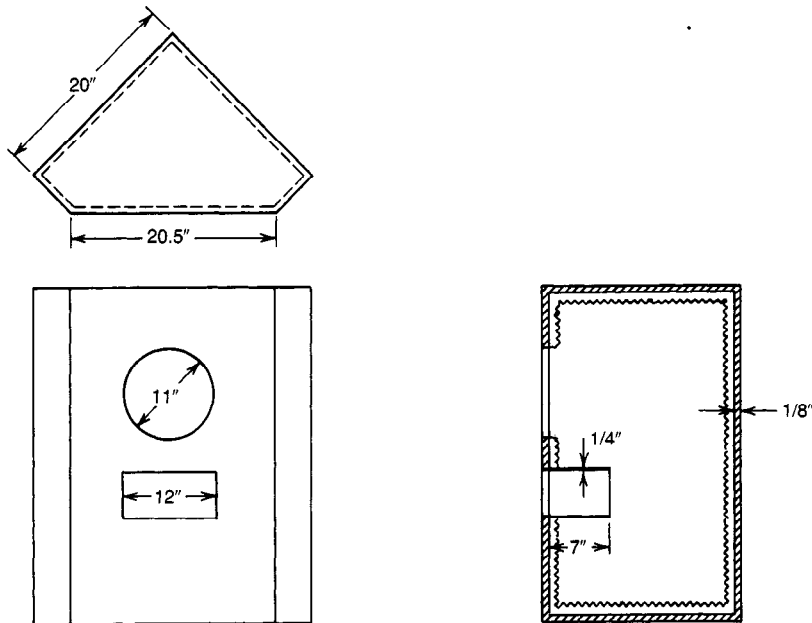


Fig. 45-33. Baffle con escape de bajos diedro.

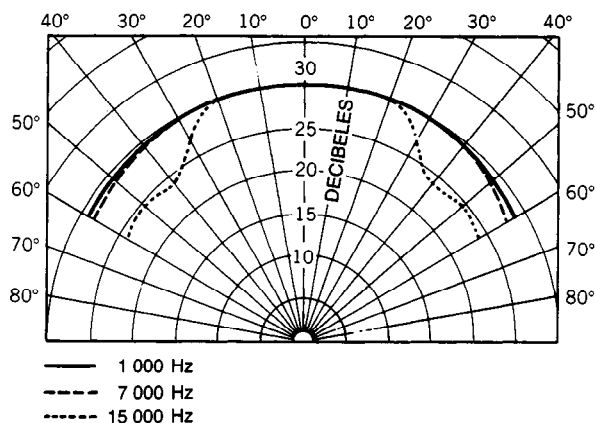


Fig. 45-34. Curvas polares de un altavoz típico de alta calidad y amplio intervalo en un mueble acústico. Cortesía de Audio Cyclopedia.

tavoz y la correlación de éstas con las que el oyente interpreta como "correctas". También entra en la controversia el tipo de micrófono de medición que debe utilizarse, ya que ningún micrófono omnidireccional es capaz de tener respuestas perfectamente planas sobre y fuera del eje. Cada diseñador de altavoces ha llegado necesariamente a su propio conjunto de curvas de corrección estándar que aparentemente se correlacionan mejor con la dispersión de los altavoces, el tipo de micrófono utilizado y las características de reflexión de la sala de audición. A pesar de los avances recientes en las técnicas de medición, incluyendo la espectrometría de retardo y el análisis rápido de Fourier, aún no es viable una determinación científica inequívoca de la respuesta de frecuencia "correcta".

De hecho, las técnicas de medición por computadora permiten actualmente aprender más acerca de los altavoces de lo que es posible interpretar por completo. Algunas de las mediciones más factibles están relacionadas con el análisis de los nodos de vibración de los altavoces, la curva energía-tiempo de un sistema (ETC, de *energy-time curve*) y la respuesta de fase. Aunque incluso altavoces de "buen sonido" miden de forma imperfecta en varias escalas empíricas, aún no se han definido ventanas de aceptación de audibilidad para ciertos parámetros, como la ETC, el error de fase y la distorsión IM.

BIBLIOGRAFÍA

REFERENCIAS BIBLIOGRÁFICAS

1. A. N. Thiele, "Loudspeakers in Vented Boxes" Parte I, *JAES*.
2. A. N. Thiele, "Loudspeakers in Vented Boxes" Parte II, *JAES*.
3. R. H. Small, "Closed-Box Loudspeaker Systems" Parte I, *JAES*.
4. R. H. Small, "Closed-Box Loudspeaker Systems" Parte II, *JAES*.

LIBROS Y REVISTAS

- B. G. Belkin, "The Need for Transient Distortion Measurements in Loudspeakers", *Sov. Phys.-Acoust.* **18**(2).
- L. L. Beranek, *Acoustics*, McGraw-Hill, Nueva York.
- F. H. Brittain, "Metal Cone Loudspeaker", *Wireless World*.
- A. D. Broadhurst, "Loudspeaker Enclosure to Simulate an Infinite Baffle", *Acustica* **39**.
- C. D. Capetanopoulos, "Measurement of the Directivity Characteristics of Loudspeakers and Microphones in a Reverberant Enclosure", *Trans. Audio Electroacoust.* **AU-20**(2).
- M. Collems, *High Performance Loudspeakers*, Wiley, Nueva York.
- E. de Boer, "Theory of Motional Feedback", *IRE Trans. Audio* **AU-9**(1).
- C. A. Ewaskio y O. K. Mawardi, "Electroacoustic Phase Shift in Loudspeakers", *J. Acoust. Soc. Am.* **22**(4).
- F. J. M. Frankort, "Vibration Patterns and Radiation Behaviour of Loudspeaker Cones", *Philips Tech. Rev.* **36**(1).
- C. L. Gilford, "The Acoustic Design of Talk Studios and Listening Rooms", *Proc. IEE* **106**, Parte B(27).
- C. R. Hanna y J. Slepian, "The Function and Design of Horns for Loudspeakers-Reprint", *J. Audio Eng. Soc.* **25**(9). "Discussion: The Function and Design of Horns for Loudspeakers". Reimpreso de *Trans. AIEE*, 1924, *J. Audio Eng. Soc.* **26**(3).
- H. D. Harwood, "New BBC Monitoring Loudspeaker: Design of the Low-Frequency Unit", *Wireless World*.
- H. D. Harwood, "Loudspeakers Developments", *Br. Acoust. Soc. Electroacoust. Air Water*.
- H. D. Harwood, "Testing High-Quality Loudspeakers", Parte II, *Audio*.

- H. D. Harwood, "Some Factors in Loudspeaker Quality", *Wireless World*.
- R. C. Heyser, "Acoustical Measurements by Time Delay Spectrometry", *J. Audio Eng. Soc.* **15**(4).
- H. W. Holdaway, "Design of Velocity-Feedback Transducer Systems for Stable Low-Frequency Behaviour", *IEEE Trans. Audio* **AU-11**(5).
- F. V. Hunt, *Electroacoustics*, Wiley, Nueva York.
- J. A. Klaassen y S. H. de Koning, "Motional Feedback with Loudspeakers", *Philips Tech. Rev.* **29**(5).
- S. Klein, "The Ionophone", *Onde Electrique* **32**.
- P. W. Klipsch, "A Low-Frequency Horn of Small Dimensions", *J. Acoust. Soc. Am.* **13**.
- S. H. Linkwitz, "Loudspeaker System Design", *Wireless World*.
- N. W. McLachlan, *Loudspeakers*, Oxford University Press, Gran Bretaña, reprinted Dover, Nueva York.
- H. F. Olson, *Elements of Acoustical Engineering*, D. Van Nostrand, Nueva York.
- H. F. Olson, *Acoustical Engineering*, D. Van Nostrand, Princeton.
- G. Rajkai, "Investigation of the Distortion of Dynamic Loudspeakers at Low Frequencies", *Seventh International Congress on Acoustics, Acoustical Commission of Hungarian Academy of Sciences, Budapest*.
- C. W. Rice y E. W. Kellogg, "Notes on the Development of a New Type of Hornless Loudspeaker", *J. Am. Inst. Elec. Engrs.* **44**(9).
- D. E. L. Shorter, "Loudspeaker Transient Response—Its Measurement and Graphical Representation", *BBC Q.* **1**.
- D. E. L. Shorter, "A Survey of Performance Criteria and Design Considerations for High-Quality Monitoring Loudspeakers", *IEE Informe n.º 2604*.
- R. H. Small, "Efficiency of Direct-Radiator Loudspeaker Systems", *J. Audio Eng. Soc.* **19**(10).
- R. H. Small, "Performance Limitations and Synthesis of Direct-Radiator Loudspeaker Systems", *Proc. IREE* **34**(8).
- W. R. Stroh, "Phase Shift in Electroacoustic Transducers", *Acoust. Res. Lab., Harvard Univ. Tech. Mem.* **42**.
- A. N. Thiele, "Loudspeakers, Enclosures and Equalisers", *Proc. IREE* **34**(11).
- P. J. Walker, "Wide Range Electrostatic Loudspeakers", *Wireless World*.
- R. E. Werner, "Loudspeakers and Negative Impedances", *IRE Trans. Audio* **AU-6**(4).

CAPÍTULO 46

Grabación y reproducción de vídeo

RCA Corporation

Consumer Electronics Technical Training
Indianapolis, Indiana

46.1 Fundamentos de la grabación en cinta de vídeo

- 46.1.1 Introducción
- 46.1.2 Sistema mecánico
- 46.1.3 Sistema de grabación
- 46.1.4 Sistema de reproducción
- 46.1.5 Sistema de color
- 46.1.6 Sistema de control
- 46.1.7 Grabación acimutal
- 46.1.8 Resumen

46.2 Reproductor de discos de vídeo

- 46.2.1 Introducción
- 46.2.2 Controles del usuario
- 46.2.3 Teoría de la transducción por capacitancia
- 46.2.4 Señales del disco de vídeo
- 46.2.5 Operación de la cubierta portadora
- 46.2.6 Descripción general del sistema

46.1 FUNDAMENTOS DE LA GRABACIÓN EN CINTA DE VÍDEO*

46.1.1 Introducción

La mayoría de los tecnólogos están más o menos familiarizados con la forma en la que se graba información de audio en cinta magnética. El objetivo de la grabación de vídeo es esencialmente el mismo: convertir la imagen e información de sincronización en energía eléctrica y almacenarla en una cinta en la forma de campos magnéticos variantes impresos en una cubierta metálica sobre la cinta. Parecería posible grabar vídeo con la misma técnica que se aplica en la grabación de audio, hasta que se consideran las características diferentes de las dos señales y el mecanismo de grabación.

El objetivo con el vídeo es grabar y reproducir frecuencias de vídeo. Asimismo, se deben considerar otros criterios específicos durante el diseño de un sistema. Entre éstos se cuentan básicamente:

1. Buena resolución de vídeo
2. Tiempo de grabación y reproducción aceptable (uso mínimo de la cinta).
3. Operación simplificada (suponiendo que se emplea como un producto de entretenimiento doméstico).
4. Mecanismos confiables y exactos.
5. Capacidad de reproducción a través de un receptor de televisión.

En el campo del audio, se manejan señales en una escala de frecuencia de cerca de 20 a 20 000 Hz. En el vídeo, la escala de frecuencia varía de 30 Hz a 4 MHz (fig. 46-1). Esta diferencia plantea varios problemas cuando se consideran las características de una cabeza de grabación y el sistema asociado. Se descubrirá que el hueco o entrehierro de la cabeza impone la limitación (influencia) más grande en un sistema para la grabación del espectro de vídeo.

En la figura 46-2 se muestra que la relación entre el ancho del entrehierro de la cabeza de grabación y la longitud de la onda en la cinta se refleja en la salida de voltaje, e indica que la salida es máxima cuando el entrehierro es un medio de la longitud de onda y que existe un punto más allá del cual el sistema es inefectivo. Esto se vuelve más significativo cuando se aplica la fórmula:

$$\lambda = \frac{V}{F}$$

donde λ = longitud de onda de grabación, pulg/ciclo

V = velocidad de la cinta, pulg/s

F = frecuencia de la señal, ciclos/s

Si se considera una simulación de audio hipotética con velocidad de la cinta de 15 pulg/s y respuesta de frecuencia deseada de 20 000 ciclos/s, se tiene que $\lambda = 15/20\,000 = 0.00075$, o bien, $\lambda = 0.75$ mpulg. En

este caso, un sistema con entrehierro de la cabeza de 0.75 mpulg no generaría salida a 20 000 ciclos/s, mientras que un sistema con entrehierro de un medio de esa cantidad (0.375 mpulg) produciría una salida máxima a esta frecuencia. En las grabadoras de cinta de audio es común un ancho de entrehierro de 0.3 mpulg.

Si se aplica este mismo principio a una señal de vídeo con un valor conservador de 3.0 MHz, se encuentra que es necesaria una cabeza de grabación con ancho de entrehierro de 0.0025 mpulg a 15 pulg/s para lograr una salida máxima a 3.0 MHz.

Sin embargo, si se utiliza una cabeza de 0.3 mpulg y se incrementa la velocidad de la cinta lo suficiente para grabar 3 MHz (lo cual es posible), se emplearían 900 pulg de cinta por segundo, de modo que la cantidad necesaria de cinta y su manejo pasan a ser un problema de consideración.

Una mejor manera de incrementar la velocidad de cabeza a cinta (y resolver parte del problema) consiste en mover la cabeza y la cinta, en vez de desplazar sólo la cinta por una cabeza estacionaria. Se emplea un sistema de cabezas giratorias para incrementar la velocidad relativa entre la cabeza y la cinta. Este sistema, junto con un entrehierro de cabeza seleccionado (reducido), hace posible grabar y reproducir las altas frecuencias de vídeo.

Esta técnica puede proporcionar la velocidad de cabeza a cinta necesaria y seguir permitiendo un tiempo de grabación aceptable sin consumir una cantidad de cinta impráctica.

Sería deseable grabar sólo las frecuencias más altas, pero recuérdese que se necesita grabar un amplio intervalo de frecuencias. Este problema se hace evidente cuando se examinan la curva teórica de 6 dB por octava y las características de salida, como se presentan en la figura 46-3. Puesto que se experimenta un ascenso de 6 dB en la salida por cada octava y el intervalo de vídeo abarca cerca de 18 octavas, se obtiene una diferencia de 100 dB en la salida entre los extremos bajo y alto de dicho intervalo. Éste es un intervalo demasiado amplio para ajustarse, y simplemente incrementando la velocidad de la cinta no es posible alcanzar el objetivo propuesto. Es evidente que se necesita algún otro método.

Una manera de resolver el problema de salida consiste en desplazar hacia arriba del espectro de frecuencias el ancho de banda de 4 MHz necesario, seleccionando un intervalo de salida aceptable en ese ancho de banda. Por ejemplo, de 20 Hz a 4 MHz representa 18 octavas, pero de 4 a 8 MHz sigue siendo una dispersión de 4 MHz y sólo representa una octava y un diferencial de salida de 6 dB. (Una octava se considera una relación de frecuencias de 2:1. Aquí, si se elige de 20 a 40 Hz como la octava uno, de 40 a 80 Hz se considera la siguiente octava, de 80 a 160 Hz la octava tres y la relación [octava] 2:1 amplía hacia arriba el espectro de frecuencia cada vez que ésta se duplica.)

El método más utilizado en la actualidad consiste en imprimir una portadora en la cinta y modular su frecuencia con la señal de vídeo (fig. 46-4). Las portadoras suelen estar en el intervalo de 2 MHz. La que se ilustra en la figura 46-4 utiliza una frecuencia de porta-

*Copyright por RCA Consumer Electronics.

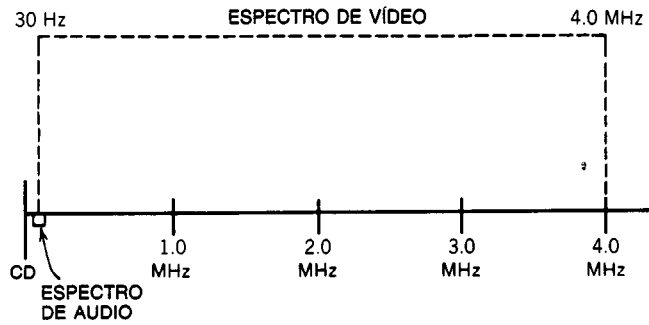


Fig. 46-1. Espectro de frecuencias de vídeo.

dora de 3.4 MHz; en este caso se producen crestas de sincronización a 3.4 MHz y crestas de blanco a 4.4 MHz. Esa figura ilustra asimismo un sistema de grabación en color bastante común. La información de croma se convierte a una banda en el área de 629 kHz del espectro antes de la grabación.

El sistema básico que se acaba de analizar posibilitó la resolución del problema del diferencial de salida para una respuesta de frecuencia en el intervalo de vídeo de los 4 MHz.

46.1.2 Sistema mecánico

La primera grabadora de cinta de vídeo práctica se produjo para la industria de la televisión comercial. Se diseñó para cumplir especificaciones muy estrictas, y después de algunos tropiezos se convirtió en un sistema muy satisfactorio que se sigue empleando en la actualidad. La mayor parte de la programación televisiva proviene de este tipo de aparato. El sistema se vale de cuatro cabezas de grabación montadas en una rueda que gira de manera que las cabezas barren las pistas de la cinta (similares a las trayectorias simplificadas que se muestran en la fig. 46-5). Utiliza una cinta de 2 pulg de ancho, que avanza a una velocidad de 14 pulg/s. Si bien este sistema realiza un trabajo excelente, es demasiado complejo y costoso para emplearse en el hogar.

El aparato de cuatro cabezas graba unas 16 líneas de vídeo en cada pista, y se requieren 16 pistas para un

campo completo de imagen de televisión (poco más de 262 líneas). Para el hogar, sería mucho más deseable un método que grabara un campo completo, en una pista. Sin embargo, sería insatisfactorio reducir la velocidad de la rueda de cabezas, ya que la velocidad de cabeza a cinta sería demasiado baja para producir la respuesta de frecuencia adecuada.

De esta forma se produjo el sistema de barrido helicoidal para uso institucional y doméstico. En vez de tener cuatro cabezas que recorren la cinta en ángulos casi rectos (o perpendiculares), la cinta se enrolla en torno a un tambor rotatorio que contiene una o dos cabezas y hace su recorrido perpendicularmente a la dirección de avance de la cinta (fig. 46-6). Mediante la selección del diámetro del cilindro de cabezas, grado de arrollamiento de la cinta, velocidad de la cinta y velocidad del cilindro, es posible grabar una pista lo suficientemente larga (a lo ancho de la cinta formando un ángulo) para contener un campo de televisión completo y conservar la velocidad de cabeza a cinta adecuada. Otro concepto de barrido helicoidal se valía de un arrollamiento de cinta "completo" en torno al tambor (fig. 46-7) y utilizaba una cabeza de grabación; en la mayor parte de los aparatos subsiguientes se emplea un semiarrollamiento con dos cabezas (fig. 46-8). Se aplican varios métodos para formar la trayectoria o el seguimiento (o ambos) de la cinta de cabeza a tambor. Los carretes de suministro y receptor se pueden colocar a diferentes alturas respecto al nivel del tam-

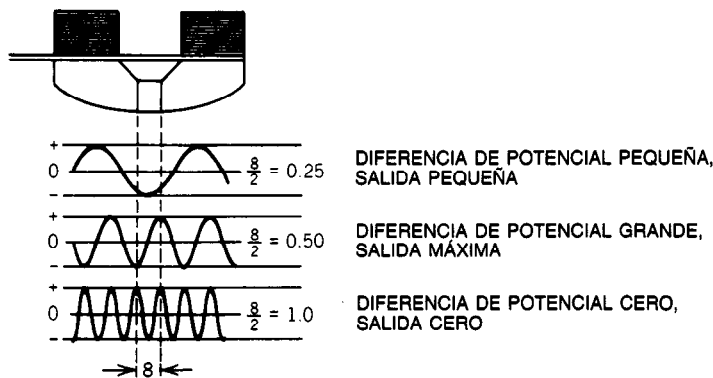


Fig. 46-2. Relación de entrehierro y longitud de onda.

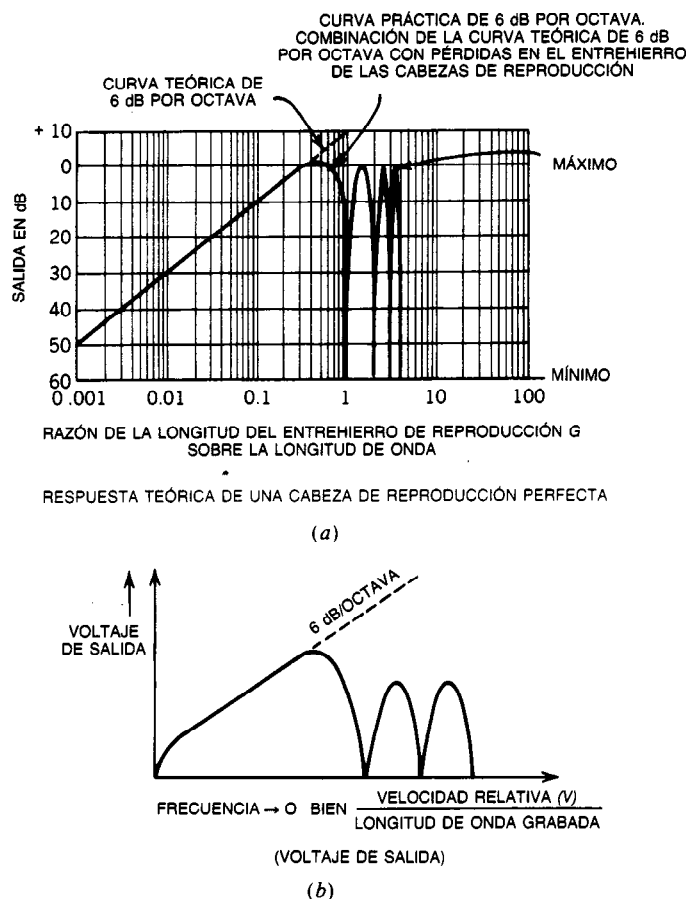


Fig. 46-3. a) Respuesta teórica de una cabeza. b) Salida simplificada de una pendiente de 6 dB por octava.

bor; el tambor cilíndrico de cabezas en sí puede inclinarse (montarse en ángulo), con los carretes de cinta paralelos. Esta disposición del tambor inclinado es probablemente la más común, ya que permite la carga en paralelo y otras ventajas.

En la figura 46-9 se ilustra una disposición común del tambor cilíndrico de cabezas. Las cabezas de grabación se separan 180° y sobresalen ligeramente de la superficie del tambor. La energía se transmite hacia (o desde) las cabezas mediante el uso de un transformador rotatorio como parte integral del montaje del tambor. La rueda gira a 30 rps (Hz), de manera que cada cabeza graba un campo completo, y una rotación completa de la rueda de cabezas graba un cuadro completo (2 campos) de información de televisión (525 líneas).

La grabación resultante se coloca en la cinta como se ilustra en la figura 46-10a. Cada pista comprende un campo de televisión, que consta de 262 1/2 líneas horizontales, incluyendo toda la información asociada, tal como las señales de luminancia, señales de croma, pulsos de sincronización cromática, pulsos de sincronización y cualquier otro material que pueda haberse transmitido en la señal con excepción del audio. La

portadora de sonido se desmodula en el sistema y se graba en la cinta como audio. La pista de control (fig. 46-10b) consta de pulsos generados y grabados por el sistema VCR (de *video cassette recorder*, grabadora de casetes de vídeo o videograbadora), que se utilizan para mantener las cabezas de vídeo debidamente alineadas con la grabación en la cinta durante la reproducción. Las diversas cabezas se colocan como se muestra en la figura 46-11. Como sucede en la grabación en cinta de audio, la cinta pasa sobre una cabeza de borrado total para eliminar cualquier magnetismo residual. La cinta pasa entonces por las cabezas de vídeo rotatorias, donde se graba la información de imagen, y después pasa por el montaje de cabezas de audio (dos), donde se graba la información de las pistas de audio y control. En el modo de reproducción, se utilizan las mismas cabezas de vídeo, audio y control para recuperar las señales de la cinta.

En una VCR helicoidal típica se utilizan tres motores: uno para impulsar el cilindro de cabezas, otro para producir el movimiento hacia delante y atrás de la cinta vía el eje de transporte y uno más para cargar y descargar la cinta en el mecanismo.

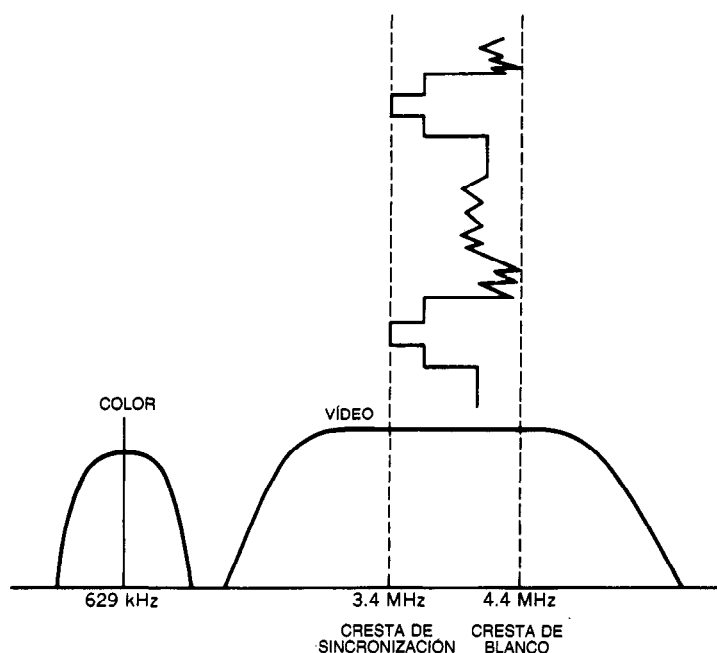


Fig. 46-4. Espectro de vídeo típico grabado.

Cilindro de cabezas. Se acopla directamente a su motor (fig. 46-9). Como la velocidad de rotación y la posición de las cabezas (respecto a las pistas de reproducción de la cinta) son críticas, el motor del cilindro es controlado por un servosistema.

Motor del eje de transporte. También es controlado por un servosistema que hace que la cinta recorra las cabezas de vídeo a una velocidad controlada con exactitud. Como sucede con el cilindro de cabezas, la velocidad de rotación y la posición del eje de transporte son críticas.

La mayor parte de los motores del eje de transporte que se utilizan en la actualidad son de accionamiento directo (DD), como en el motor del cilindro de cabezas (fig. 46-12).

Motor de carga y descarga. Simplemente proporciona la energía mecánica necesaria para operar el mecanismo de carga y descarga, y su importancia relativa es menor.

En una VCR helicoidal moderna pueden emplearse asimismo tres o más solenoides controlados electrónicamente para objetivos tales como la activación del rodillo opresor del transporte o los mecanismos de avance y rebobinado rápidos.

Un diagrama de bloques simplificado del sistema electrónico de una VCR de dos cabezas (fig. 46-13) indica algunas de las complejidades que intervienen en el cumplimiento de este objetivo. Como se mencionó antes, las cabezas sirven a un doble fin: la grabación y la reproducción.

46.1.3 Sistema de grabación

La parte principal de la VCR son las cabezas de grabación y reproducción de vídeo. Como ya se dijo, la cabeza debe ser capaz de grabar y reproducir señales en la región de los 4 MHz. A pesar de la mayor velocidad de cabeza a cinta que produce el sistema de barrido helicoidal, la capacidad de reproducir esas frecuencias requiere un entrehierro más bien pequeño.

Las cabezas de vídeo modernas se hacen de material ferroso conformado (fig. 46-14). Las cabezas típicas se elaboran con material de unas 10 mpulg de espesor y tienen un entrehierro de 40 a 50 μm .

La información de vídeo se imprime en la cinta en la forma de una portadora de onda cuadrada modulada

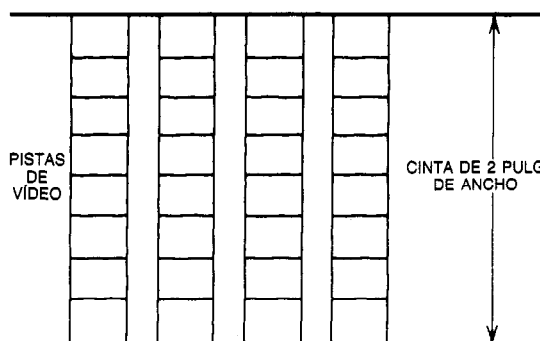


Fig. 46-5. Trayectoria de pistas de una grabadora de cuatro cabezas.

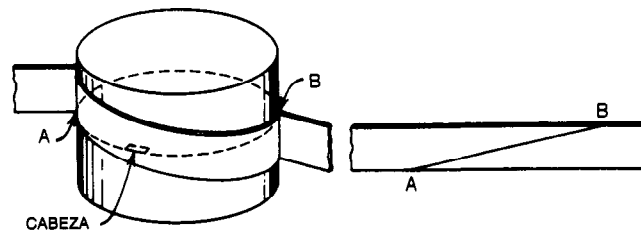


Fig. 46-6. Pista de grabación de barrido helicoidal.

en frecuencia. La figura 46-15 ilustra la forma en que se vería la impresión magnética cuando la portadora es modulada por un pulso de sincronización horizontal. La frecuencia a la cual ocurren los pulsos es función de la intensidad de la información de vídeo. Los blancos están en la más alta frecuencia (4.4 MHz), y las crestas de sincronización (negros) están en la más baja frecuencia (3.4 MHz), como se ilustra en la figura 46-14. El espacio de tiempo que dura un ritmo repetitivo dado es función del ancho horizontal (o de la frecuencia) de la información de vídeo.

La figura 46-16 es un diagrama de bloques simplificado del procesamiento al cual se somete la información de vídeo a fin de prepararla para la grabación.

Se aplica una señal de vídeo estándar al amplificador de entrada. La ganancia de esta fase se controla para garantizar una salida estable. Después, la señal se envía a un filtro pasabajas para suprimir cualquier información por encima de los 3.8 MHz, y el circuito de fijación de nivel restituye el nivel de CD de la señal, refiriéndola a crestas de sincronización. Después, un circuito de preacentuación o preénfasis acentúa la respuesta de alta frecuencia, y un circuito recortador limita la cantidad de sobrecorrección generada en ambos extremos del espectro. La señal se aplica a un modulador FM y después a un amplificador de grabación que acciona (en paralelo) las cabezas de grabación a través de transformadores rotatorios.

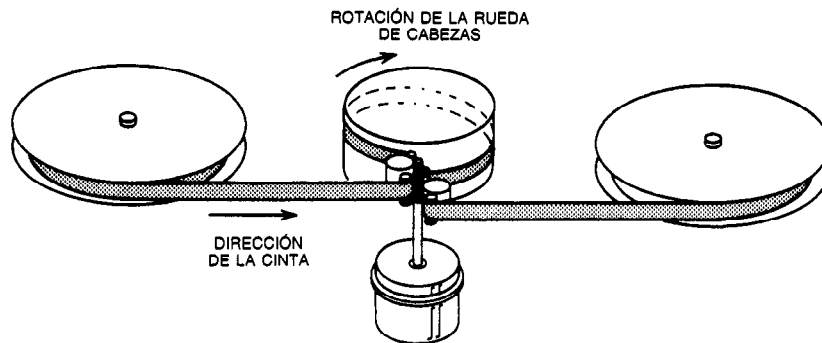


Fig. 46-7. Esquema que ilustra el concepto de barrido helicoidal con arrollamiento completo.

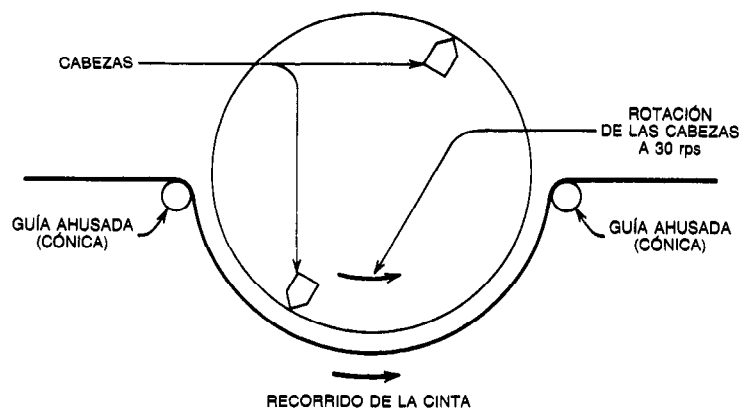


Fig. 46-8. Esquema de un barrido helicoidal con medio arrollamiento con dos cabezas.

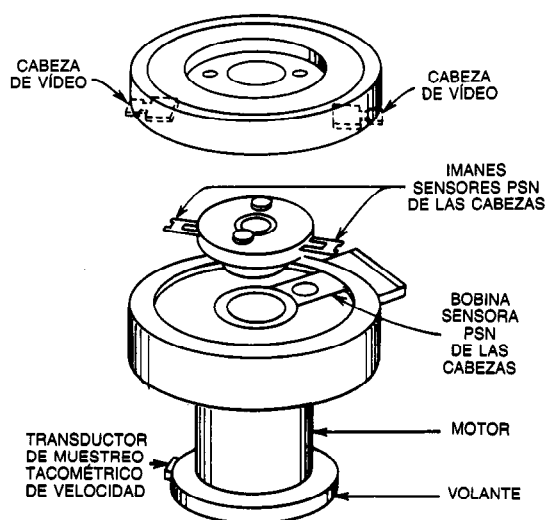


Fig. 46-9. Vista de un tambor típico de cabezas.

46.1.4 Sistema de reproducción

En la reproducción, el proceso esencialmente se invierte, ya que ahora las cabezas recogen la información grabada en la cinta y la suministran a los circuitos de reproducción. Sin embargo, esta vez las cabezas no están en paralelo. Cada cabeza se conecta a través de su propio transformador rotatorio a su preamplificador. Estos amplificadores se encienden y apagan (conmutan) a razón de 60 Hz, de tal forma que sólo la cabeza que se halla en contacto con la cinta esté en el circuito en un instante dado. Esto se hace porque la

cabeza que no está en contacto con la cinta introducirá ruido indeseable en la imagen. Las salidas de los preamplificadores se conectan entre sí, y entonces las dos señales se vuelven una sola (fig. 46-17).

La salida combinada de las dos cabezas de grabación se aplica a una serie de limitadores, donde se elimina cualquier variación de amplitud generada por ruido, señales extrañas o variaciones en el contacto. Después se aplica la señal limpia a los circuitos desmoduladores, donde se vuelve a convertir en vídeo.

Uno de los problemas que se encuentran en la reproducción de vídeo ocurre cuando, por alguna razón, la cabeza deja de hacer contacto con la cubierta de óxido metálico de la cinta. Esto puede deberse a polvo en la cinta o a defectos mínimos en la cubierta. El fenómeno resultante se conoce como exclusión o caída de señal, y si se pasa por alto puede generar parpadeo inaceptable en la imagen. Para compensar esta situación, se utiliza un circuito llamado compensador de exclusión (o de caída de señal).

La figura 46-18 muestra el diagrama de bloques de un compensador de exclusión típico. Durante la reproducción, la información de vídeo sigue la trayectoria marcada como vídeo directo. La señal se aplica asimismo a un circuito de retardo, que almacena una línea de vídeo completa. Sin embargo, la información demorada no entra en la trayectoria de la señal porque su entrada al selector es apagada por el rectificador de FM y su circuito de conmutación.

En caso de que la portadora de FM caiga por debajo de un nivel predeterminado, el voltaje generado por el rectificador de FM abrirá el interruptor y permitirá que la información almacenada en el circuito de retardo entre en la trayectoria de la señal. El resultado es que si hay una caída de señal de la portadora, el compensa-

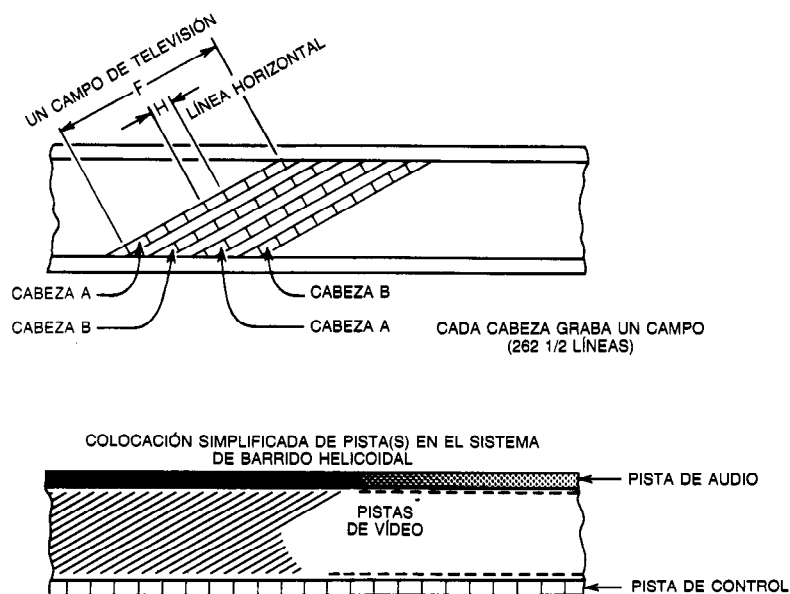


Fig. 46-10. Por lo general se graban tres pistas: audio, vídeo y control.

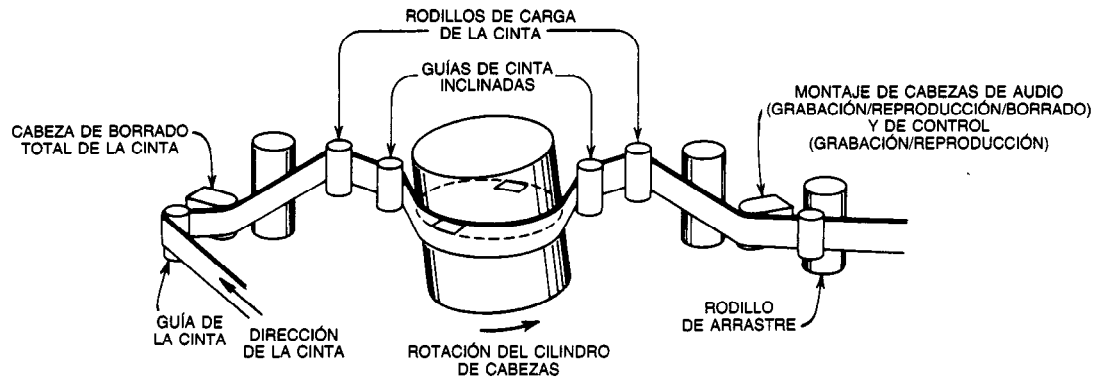


Fig. 46-11. Disposición típica de las cabezas en una VCR de barrido helicoidal.

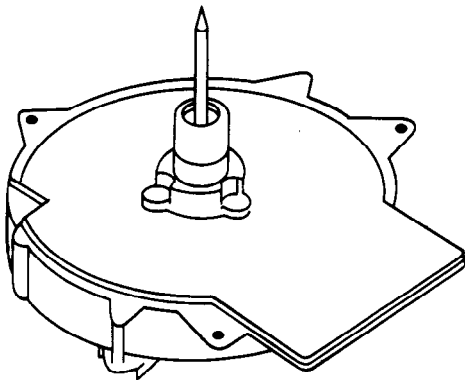


Fig. 46-12. Motor de accionamiento directo (DD) del eje de transporte.

El cabezal introduce la información que apareció en la línea anterior y continúa la imagen, sin que la interrupción sea percibida por el ojo humano. En la práctica, la compensación puede proceder en tres o cuatro líneas, debido a la trayectoria de recirculación. En ese punto, la calidad de la imagen puede desmerecer en forma perceptible.

46.1.5 Sistema de color

En la mayor parte de los sistemas de grabación de vídeo de barrido helicoidal se aplica algún método para separar la información de color y la información de luminancia antes de que la señal se grabe en la cinta. Un método común consiste en convertir la croma en alguna otra frecuencia. El sistema que se representa en la figura 46-19 convierte la subportadora de 3.58 MHz a 629 kHz, donde las bandas laterales de color figuran

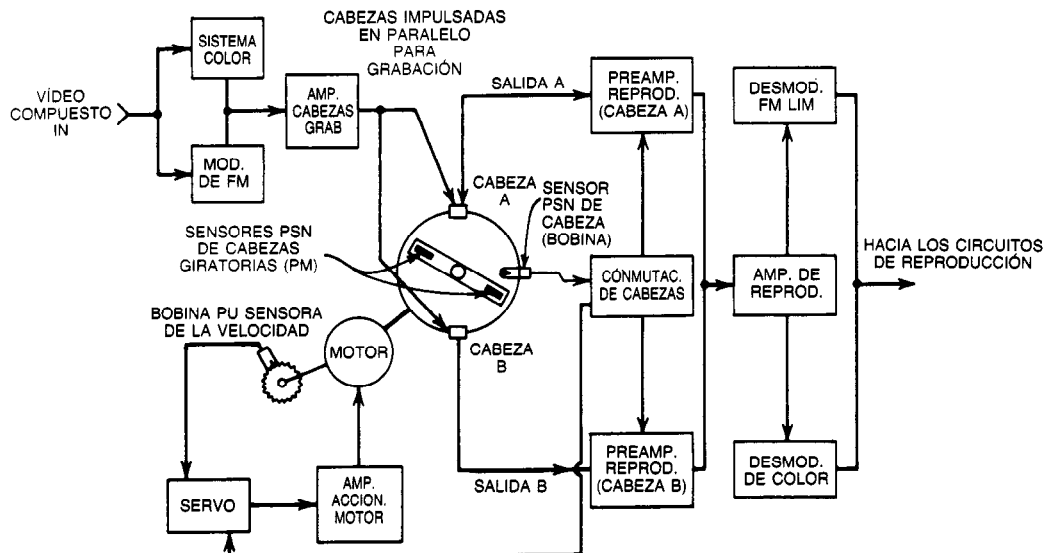


Fig. 46-13. Diagrama que ilustra la vinculación típica entre la electrónica de reproducción y la de grabación. PREAMP. = preamplificador; REPROD. = reproducción; DESMOD. = desmodulador; AMP. = amplificador; GRAB. = grabación; MOD. = modulador; ACCION. = accionamiento.

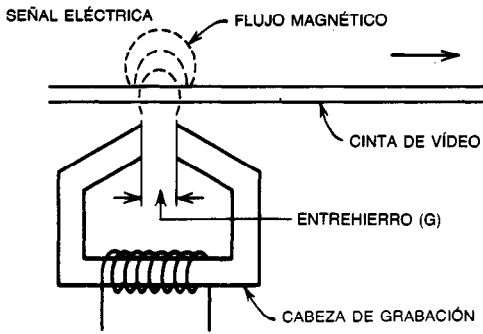


Fig. 46-14. Muestra de la forma de la cabeza de grabación y reproducción de vídeo.

en torno a ese punto. En este sistema se deben tomar medidas para limitar la excursión de baja frecuencia de la señal FM de luminancia, a fin de evitar patrones de interferencia en las partes con color de la imagen.

El proceso de conversión de la señal de 3.58 MHz a 629 kHz se ilustra en el diagrama de bloques de la figura 46-20. La croma se separa de la señal de vídeo compuesta antes de que entre en la cadena de gra-

bación de luminancia. Después, la señal de 3.58 MHz se bate (heterodina) contra un oscilador de 4.20 MHz, y la frecuencia de 629 kHz resultante, que contiene toda la información de croma, se vuelve a combinar con la señal de luminancia y se graba en la cinta.

46.1.6 Sistema de control

Para reproducir una imagen de una grabación de vídeo magnética, la cabeza de reproducción debe barrer la pista grabada con gran exactitud. Por ejemplo, en el sistema de barrido helicoidal de dos cabezas, cada línea de información grabada contiene un campo completo de la señal de televisión, y la conmutación de cabezas ocurre justo antes del borrado vertical. Por tanto, es necesario que la cabeza de reproducción haga contacto con la cinta (pista) en el mismo punto en que lo hizo la cabeza de grabación al inicio de cada pista. En la figura 46-21 se presenta lo que podría suceder en caso de que la velocidad de la cabeza y la de la cinta salieran de sincronización. Por esta causa, la unidad de impulso de la rueda de cabezas y la propia de la cinta deben ser controladas en grabación y reproducción. El control necesario de la sincronización y la colocación se logra mediante un servosistema.

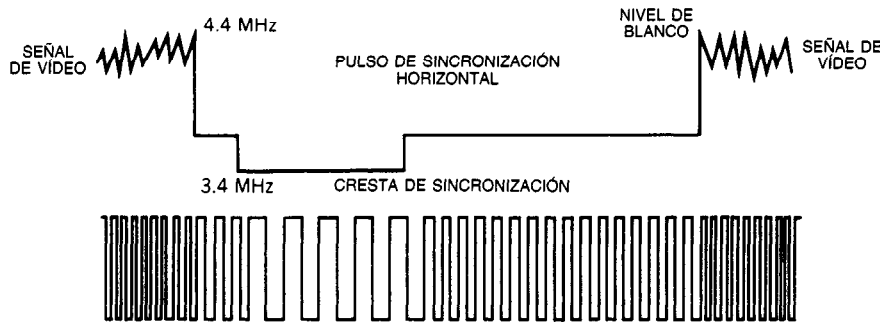


Fig. 46-15. La información de vídeo grabada en cinta es una onda cuadrada de FM.

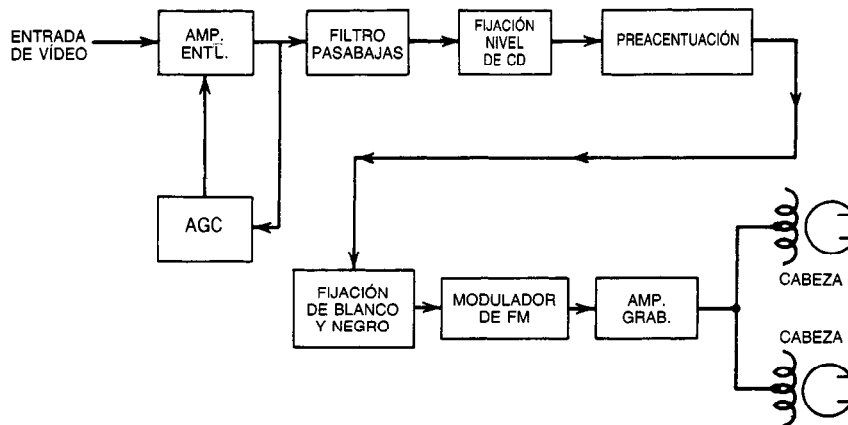


Fig. 46-16. Diagrama de bloques simplificado de una cadena de grabación de luminancia. AMP. ENT. = amplificador de entrada; AMP. GRAB. = amplificador de grabación.

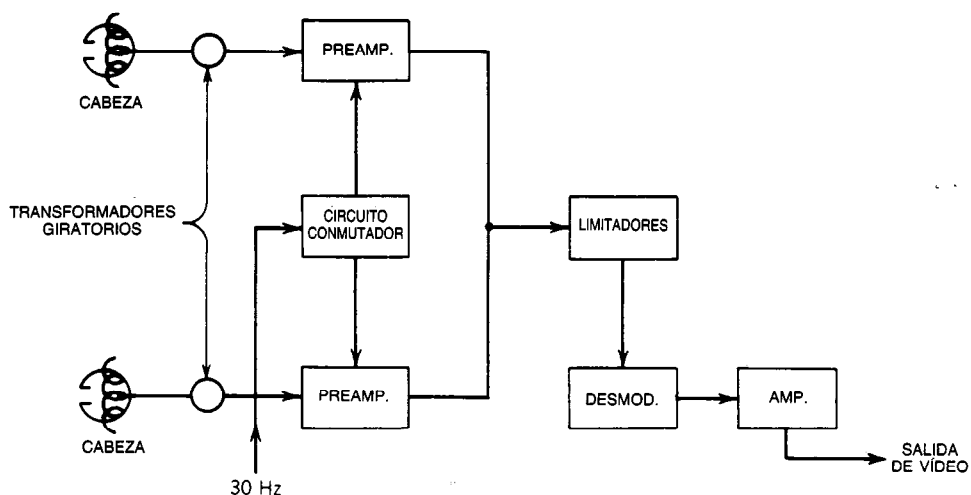


Fig. 46-17. Cadena de reproducción de luminancia simplificada. PREAMP. = preamplificador; DESMOD. = desmodulador; AMP. = amplificador.

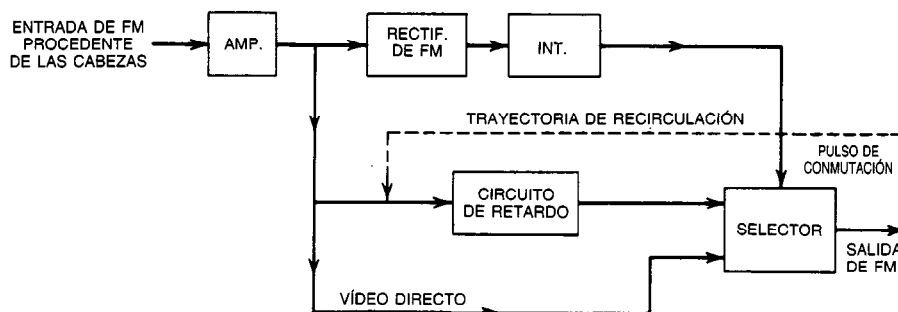


Fig. 46-18. Diagrama de bloques de un compensador de exclusión. AMP. = amplificador; RECTIF. = rectificador; INT. = interruptor.

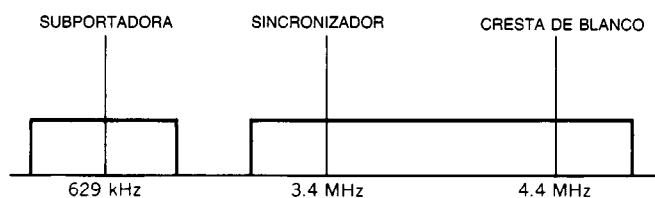


Fig. 46-19. Frecuencias de portadoras de color sobre cinta.

Un servosistema o servomecanismo es un dispositivo automático para controlar y corregir el funcionamiento de un mecanismo. Para realizar su función, un servo necesita cuando menos dos tipos de información:

1. Una entrada procedente del aparato a partir de la cual pueda determinar su desempeño (retroalimentación).
2. Una fuente (de referencia) constante para determinar el grado al cual se desvía la máquina del funcionamiento deseado.

En el modo de grabación, el sistema de control comienza con el pulso de sincronización vertical de la señal de vídeo que llega (fig. 46-22); el pulso se procesa y se convierte en la entrada de referencia de un circuito de comparación. La otra entrada del circuito de comparación recibe una señal de retroalimentación generada desde el motor de accionamiento. El circuito compara las dos entradas y produce una salida que es representativa de la diferencia entre las dos. Esta salida se utiliza para controlar el motor de accionamiento y, por tanto, cualquier cambio en la relación provocará

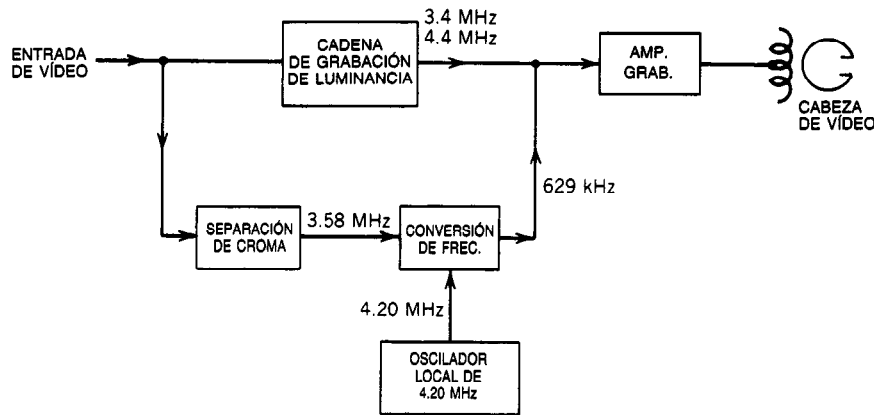


Fig. 46-20. Principio de grabación por el método de subportadora convertida. AMP. GRAB. = amplificador de grabación; FREC. = frecuencias.

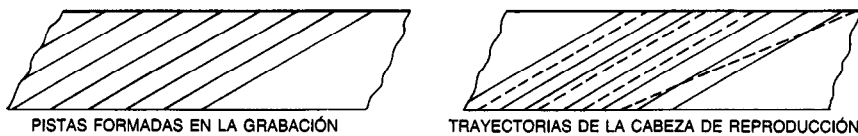


Fig. 46-21. Seguimiento en una VCR.

un cambio correctivo en el accionamiento. Esta disposición podría utilizarse para controlar la rueda de cabezas o bien el eje de transporte de cinta. En uno u otro caso, la otra unidad de accionamiento se encontraría a una velocidad constante predeterminada.

Al mismo tiempo que se graba la señal y el pulso de sincronización vertical controla un motor de accionamiento, se aplica ese mismo pulso a un divisor de frecuencia que produce pulsos de 30 Hz. Éstos se emplean para activar un circuito que graba pulsos de 30 Hz en la pista de control de la cinta (fig. 46-23).

En el modo de reproducción, el sistema de control es el mismo, salvo que la entrada de referencia son ahora los pulsos de 30 Hz que se toman de la pista de control. Esta vez el circuito de procesamiento proporciona un cierto control externo (seguimiento variable) de la sincronización de los pulsos antes de que se apliquen al circuito de comparación (fig. 46-24). En teoría, el control de seguimiento no debe ser necesario, pero en la práctica algunas veces es deseable debido al esti-

ramiento de la cinta o bien cuando se reproduce una grabación realizada en otro aparato (el seguimiento suele ser ajustable por un control del tablero frontal).

Ocurre un problema en la reproducción cuando una cabeza no explora perfectamente la pista grabada. Si la cabeza se desvía de la pista, entregará una salida reducida, pero también puede captar señales de una pista adyacente, causando interferencias. Para minimizar las interferencias se coloca una banda de protección entre las pistas. Esto es muy efectivo, pero requiere más espacio en la cinta y por consiguiente reduce la capacidad de tiempo de grabación.

Otro método de grabación, la grabación acimutal, elimina la necesidad de banda de protección.

46.1.7 Grabación acimutal

En la mayor parte de las grabaciones magnéticas se ha acostumbrado a colocar el entrehierro de la cabeza de grabación perpendicular a la dirección de rastreo. En

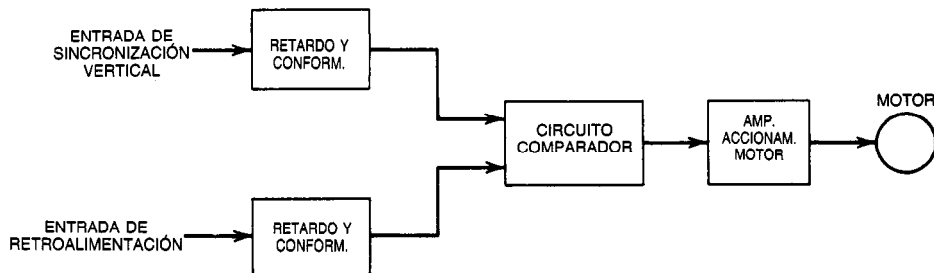


Fig. 46-22. Diagrama de bloques del servo de una VCR en modo de grabación. AMP. ACCIONAM. = amplificador de accionamiento; CONFORM. = conformación.

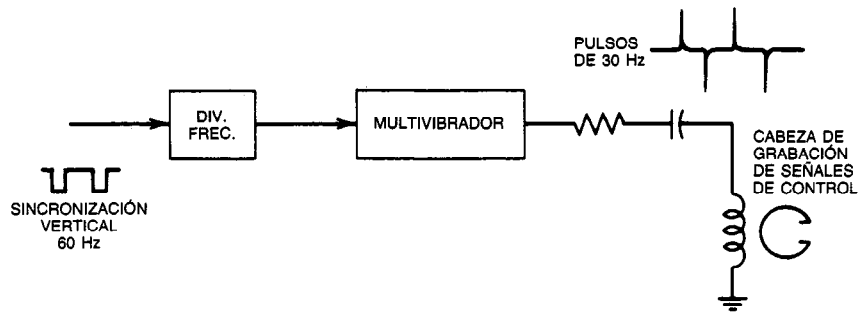


Fig. 46-23. Diagrama de bloques de una grabación de pulsos de control. DIV. FREC. = divisor de frecuencias.

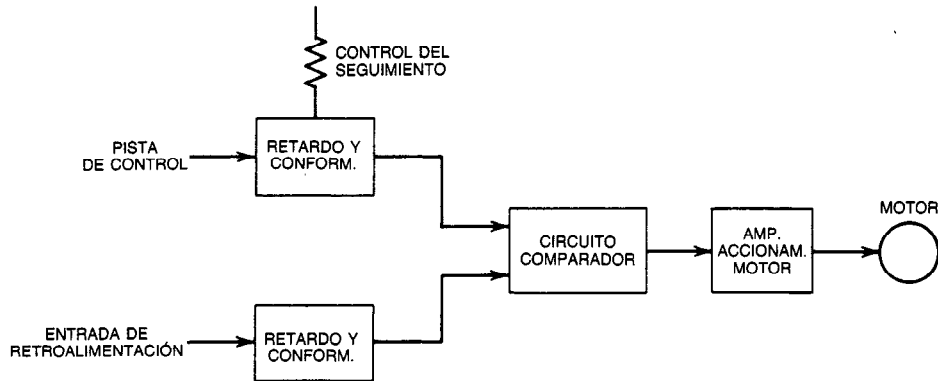


Fig. 46-24. Diagrama de bloques del servo de una VCR en modo de reproducción. AMP. ACCIONAM. = amplificador de accionamiento; CONFORM. = conformación.

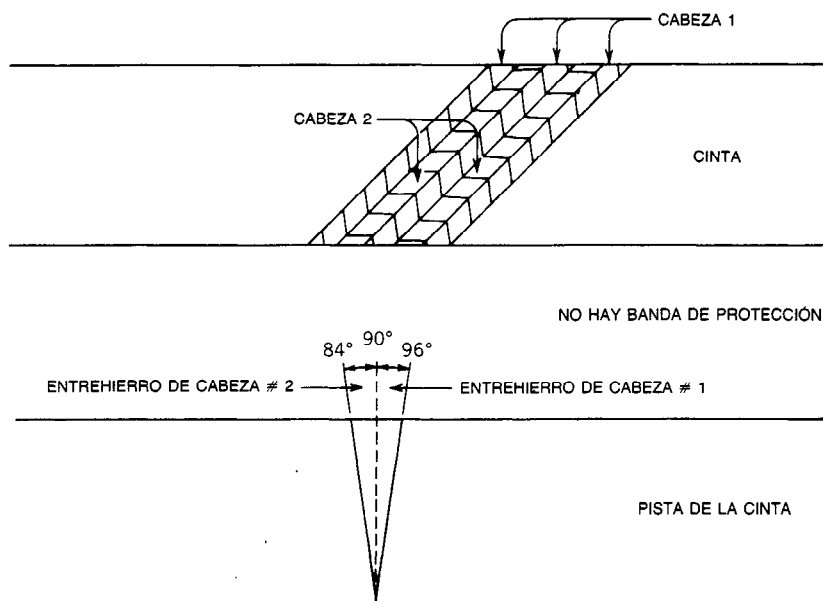


Fig. 46-25. En la técnica de grabación acimutal el entrehierro de las cabezas se coloca en distintos ángulos.

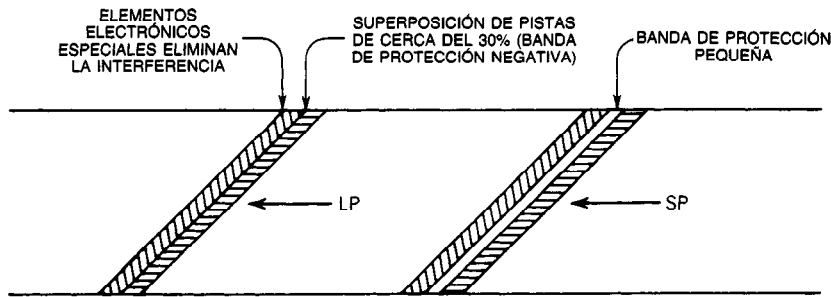


Fig. 46-26. Grabación de pistas en la VCR de RCA (Video Home System, VHS).

la reproducción, es importante que el entrehierro de la cabeza lectora esté precisamente en ese mismo ángulo, o se generarán respuestas de frecuencia y salida reducidas.

En la grabación acimutal, las dos cabezas de un sistema se colocan deliberadamente en ángulos diferentes respecto a la dirección de barrido. El sistema que se ilustra en la figura 46-25 coloca los entrehierros de las cabezas a 84 y 96°, respectivamente. Esto hace posible grabar sin banda de protección entre pistas debido a que, si una cabeza se desviara a la pista errónea, la cabeza lectora estaría a un ángulo de 12° de la señal grabada, y la interferencia sería mínima. Sin embargo, esta disposición complica el servosistema, debido a que una cabeza no sólo debe barrer una pista con exactitud, sino que también debe barrer la pista correcta en forma precisa.

En el Video Home System (VHS) se aplica la técnica de grabación acimutal, y esto hace posible grabar cuatro horas de material de programas en un casete de cinta relativamente pequeño.

La forma en la cual se graban las pistas en la cinta se ilustra en la figura 46-26. Se observará que existe una banda de protección cuando se graba en el modo de dos horas o de velocidad estándar. No obstante, cuando la máquina se opera en el modo LP (de cuatro horas), las pistas en realidad se superponen sin que se presente un efecto indeseable.

46.1.8 Resumen

La información anterior se expuso con el único objeto de presentar un breve panorama general de las grabadoras de cinta de vídeo.

Las técnicas y sistemas de grabación de vídeo han mejorado a gran velocidad. El barrido helicoidal, la grabación acimutal (con entrehierros de las cabezas a ángulos distintos), la modulación en frecuencia de información de luminancia y la grabación directa de información de color con conversión descendente (por lo general de fase rotatoria) se han convertido en factores comunes.

Diversas clavijas (jacks) de entrada y salida permiten una interfase externa con otros instrumentos, accesorios, señales o controles; las cámaras se encuentran entre los accesorios que normalmente se consideran opcionales.

LIBROS Y REVISTAS

- C. Bensinger, *The Home Video Handbook*, 3.^a ed., Howard Sams, Indianapolis.
- D. Chesire, *The Video Manual*, Van Nostrand, Nueva York.
- W. Clason, *Dictionary of Television and Video Recording*, Elsevier, Nueva York.
- J. Effrein, *Video Tape Production and Communication Techniques*, Tab Books, Blue Ridge Summit, PA.
- J. Effrein, *Home Video Yearbook*, Knowledge Industry, White Plains, NY.
- P. Kortman, *Handbook of Video Cassette Recording*, Van Nostrand, Nueva York.
- H. Kybett, *Video Tape Recorders*, 2.^a ed., Howard Sams, Indianapolis.
- P. Lanzendorf, *The Video Taping Handbook*, Harmony, Prospect, KY.
- J. D. Lenk, *Complete Guide to Video Recorder Operation and Service*, Prentice Hall, Englewood Cliffs, NJ.
- G. P. McGinty, *Video Cassette Recording, Theory and Servicing*, McGraw-Hill, Nueva York.
- R. Robinson, *The Video Primer: Equipment Production and Concepts*, Putnam, Nueva York.

46.2 REPRODUCTOR DE DISCOS DE VÍDEO*

46.2.1 Introducción

El reproductor de discos de vídeo CED (de *capacitance electronic disk*, disco electrónico de capacitancia) de RCA representa un nuevo producto que ha adquirido gran importancia por su faceta de entretenimiento doméstico. Proporciona al consumidor un medio económico de reproducir y visualizar material de programas pregrabado.

La tecnología de transducción por capacitancia, que se ha investigado y desarrollado durante muchos años en RCA, se aplica en el reproductor de discos de vídeo. El sistema de capacitancia se seleccionó para su producción después de una cuidadosa evaluación de las ventajas y desventajas de otros métodos. Se encontró

*Copyright por RCA Consumer Electronics.

que la técnica de transducción por capacitancia ofrece la mejor combinación de sencillez en la fabricación, características y rendimiento.

El reproductor de discos de vídeo de RCA presenta características como la búsqueda visual (permite el barrido del programa del disco de vídeo a 16 veces la velocidad normal) y el rápido acceso (permite la entrada rápida y fácil a un segmento específico de un programa en cualquier parte del disco de vídeo). Asimismo, se proporciona una característica "estéreo" total en modelos seleccionados para permitir al usuario reproducir discos estereofónicos o codificados en dos idiomas. También se ha incorporado un control a distancia de luz infrarroja de cinco funciones en algunos modelos, permitiendo que el usuario tenga acceso remoto a las funciones del reproductor.

El tiempo de reproducción fue asimismo un factor importante en la selección del sistema de capacitancia. El sistema CED de RCA ofrece dos horas de material de programa por disco de vídeo, o bien una hora de material de programa por cada lado.

El reproductor de discos de vídeo mismo es un instrumento muy compacto que mide 5 3/4 pulg de alto, 17 pulg de ancho y 15 1/2 pulg de fondo, y es también muy ligero (menos de 20 lb). Su consumo de energía es de apenas 35 W.

El reproductor es asimismo muy simple de operar. El disco de vídeo está almacenado en una cubierta plástica protectora o "contenedor" que se inserta en el reproductor cuando la máquina está en la posición de carga. El contenedor garantiza que el disco siempre se cargue correctamente y también que esté protegido cuando no se encuentre inserto en el reproductor. Una vez el disco se carga en la máquina, ésta se pone en el modo de reproducción.

En ese momento, una microcomputadora de control del sistema asume el control del reproductor. Al término del programa, el usuario vuelve a colocar el aparato en posición de carga y reinserta el contenedor en el reproductor para retirar el disco de vídeo. El contenedor actúa entonces como un recipiente de almacenamiento cómodo y seguro del disco.

46.2.2 Controles del usuario

Entre los controles del usuario en el reproductor de discos de vídeo de RCA se incluyen la palanca de funciones (que se encuentra en los modelos SFT075/100 y SGT100/200), la cual controla los modos de apagado, reproducción y carga. El modelo más reciente, SGT250, tiene un botón de carga y descarga y otro de encendido y apagado, que elimina la necesidad de la palanca mecánica de funciones. Los botones de avance y retroceso para acceso rápido, los de avance y retroceso para búsqueda visual y un botón de pausa están localizados asimismo en la parte frontal del reproductor.

El reproductor SGT250 Selecta Vision VideoDisc constituye otro adelanto importante en la tecnología de los reproductores de discos de vídeo. Este modelo no sólo ofrece el recurso real de la "estereofonía" que se presentó en el SGT200, sino también la comodidad

adicional del control remoto IR (infrarrojo) de todas las funciones, junto con la operación electrónica "de tacto suave".

Salvo por la adición de algunos nuevos interruptores, la configuración mecánica "básica" y su operación se mantienen esencialmente sin cambio. Estos nuevos interruptores y sus funciones respectivas se analizan con detalle más adelante.

Los circuitos básicos de procesamiento de señales del SGT250 son muy similares a los que se encuentran en los modelos SFT100 y SGT200. Las áreas principales que son nuevas y diferentes constan del control de mecanismo de "tacto suave" y los circuitos de control remoto.

La operación se mantiene muy simple y directa para el usuario, y es muy semejante a la de modelos anteriores. En muchas formas el SGT250 es más sencillo de operar que los anteriores. Por ejemplo, todo lo que debe hacer el operador para poner el aparato en el modo de reproducción es pulsar el botón de encendido, insertar el disco y después retirar el contenedor o portadora vacía. Una vez hecho esto último, la máquina se pone automáticamente en el modo de reproducción.

Tal como sucede con el mando a distancia, el usuario puede operar la unidad en avance y retroceso de acceso rápido o de búsqueda visual, y también en el modo de pausa, simplemente oprimiendo el botón adecuado en la unidad manual de control remoto IR.

En la parte posterior del reproductor de discos de vídeo está el conmutador selector de canales para la señal de salida RF. La salida RF del reproductor se puede conmutar al canal tres o cuatro con este interruptor. Asimismo, en los modelos "estereofónicos" el interruptor selector de audio, las clavijas (jacks) de salida de estéreo y la de salida de vídeo (salida de 1 V cresta a cresta) están ubicados en la parte posterior del instrumento.

El conmutador selector de audio debe estar en la posición "normal" salvo para reproducir discos con pistas de sonido en dos idiomas. Cuando se utilizan los jacks de salida estéreo, deben emplearse los cables de audio de 5 pies para conectarlos a un amplificador estéreo (no deben conectarse directamente a una serie de altavoces). Las conexiones de salida estéreo a un amplificador estereofónico deben representar una impedancia de entrada de 10 k Ω o mayor.

La palanca de la función manual (SFT100, SGT075/100/200) controla la entrada de energía al reproductor y también la operación de carga y descarga del contenedor y el disco. Conecta asimismo la entrada de antena externa directamente al conector de salida de RF cuando el reproductor está en "apagado". En los modelos más recientes, que tienen la característica de carga automática (de tacto suave), el motor de accionamiento de funciones reemplaza a la palanca manual. Este motor realiza en esencia las mismas funciones mecánicas que la palanca manual.

Para operar el reproductor manual de discos de vídeo, la palanca de función se coloca primero en la posición de carga. Esto aplica B + a los circuitos electrónicos del reproductor y también abre la puerta de en-

trada del contenedor en la parte frontal del aparato, lo que permite insertar el contenedor en el reproductor. Una vez retirada de éste el contenedor, dejando el disco y el eje en el interior, la palanca de función se coloca en la posición de reproducción. Esto hace que la aguja baje sobre el disco de vídeo, lo que permite que el reproductor comience a detectar las señales del disco, y que se generen unas imágenes en la pantalla del receptor. Ocurre una demora aproximada de seis segundos entre la colocación de la palanca de función en la posición de reproducción y la generación de señales de vídeo y de audio.

En el panel frontal del reproductor de discos de vídeo se hallan ubicados tres dispositivos indicadores (cuatro en los modelos "estéreo", SGT200/250). Uno es una pantalla LED de dos dígitos que indica el tiempo transcurrido del programa del disco en minutos. Los otros dos son los LED indicadores de los lados (caras del disco) uno y dos cuando se reproducen. Cuando se va a insertar el contenedor, la pantalla LED centellea mostrando la letra "L", lo que indica que el reproductor está ahora en el modo de carga. Cuando la palanca de función se coloca en la posición de reproducción para comenzar a reproducir un disco de vídeo, la pantalla LED muestra dos rayas (--) hasta que se recupera una señal de vídeo del disco; en ese momento la pantalla LED muestra el tiempo transcurrido del programa en minutos.

El cuarto y más reciente indicador es el LED "estéreo", que se ilumina cuando se reproduce un disco de vídeo estereofónico y el interruptor selector de audio está en la posición "normal".

Los botones de avance y retroceso de acceso rápido permiten al usuario adelantar o regresar el disco con rapidez, lo cual facilita la localización de un segmento particular del disco. Mediante el uso de los botones de acceso rápido, todo un lado del disco se puede rastrear en menos de 30 segundos. La pulsación de uno u otro botón hace que se levante la aguja del disco y que los circuitos de señales se "caigan". Esto borra la pantalla de vídeo, lo que evita que se genere ruido cuando el servo desplace rápidamente el conjunto de captación a través del disco.

Mientras opera la característica de acceso rápido, la pantalla LED proporciona una indicación aproximada (en un minuto) del tiempo de programa que ha transcurrido. En el modo de acceso rápido la aguja se levanta, así que la información del tiempo exacto que se codifica en el disco de vídeo no está disponible. El tiempo aproximado lo genera un interruptor óptico conectado al engranaje accionador del brazo de captación.

Los botones de avance y retroceso de búsqueda visual permiten al usuario explorar el programa del disco de vídeo a aproximadamente 16 veces la velocidad normal. Durante la búsqueda visual, la imagen se observa "en cámara rápida" en la pantalla. Sin embargo, la parte de audio de la señal no se reproduce en esta operación, ya que a esta velocidad las señales de audio serían incomprensibles. Al igual que la de acceso rápido, la característica de búsqueda visual opera en los sentidos de avance y retroceso, por lo que el usuario

puede explorar hacia delante o regresando una parte cualquiera del programa del disco mientras observa la señal de vídeo.

La electrónica de retracción de la aguja, que hace posibles las características de búsqueda visual, también es responsable de evitar que la aguja se "atore" en un surco. Un surco cerrado en un disco de vídeo produce un efecto similar al de un surco cerrado en un sistema de audio; a saber, las señales de vídeo y audio de ese surco se repiten una y otra vez. Como en una revolución del surco del disco están contenidos ocho campos verticales, esos ocho campos de vídeo y señales de audio se repiten. La electrónica de control del sistema, al revisar un código especial en el disco de vídeo, hace que los circuitos de retracción hagan avanzar la aguja cuando se detecta un surco cerrado.

El reproductor de discos de vídeo de RCA tiene una característica de pausa. Cuando se oprime el botón de pausa, la aguja se levanta del disco. Al mismo tiempo los circuitos electrónicos del aparato se colocan en el modo de silenciamiento o supresión ("caída de los circuitos"), lo que evita que se produzca ruido en la pantalla de vídeo. Cuando se pulsa el botón de pausa, la pantalla LED muestra la letra "P", que centellea a razón de uno por segundo. Una vez puesto en el modo de pausa, el reproductor permanece en este modo hasta que se vuelve a pulsar el botón de pausa o bien hasta que se oprimen los botones de acceso rápido o de búsqueda visual. Cuando se pulsa el botón de pausa para salir del modo de pausa, se hace bajar la aguja hacia el disco de vídeo y se vuelve a generar la señal de vídeo aproximadamente en el mismo punto del programa en que se pulsó por vez primera aquel botón. Cuando se oprimen los botones de acceso rápido o de búsqueda visual para salir del modo de pausa, el reproductor entra de inmediato al acceso rápido o hace bajar la aguja y entra a la búsqueda visual, respectivamente.

En cada cara del disco, al final del programa está grabado un código especial que hace que el reproductor pase al modo de terminación. Cuando se detecta este código, se levanta la aguja y la pantalla LED muestra una "E" centelleando a razón de una por segundo. El reproductor permanece en el modo de terminación hasta que se retira el disco o bien hasta que se pulsan los botones de retroceso, de acceso rápido o de búsqueda visual hacia atrás. Los botones de avance se desactivan en el modo de terminación. La pantalla LED centelleante en los modos de carga, pausa y terminación indica que el reproductor se mantiene en estos modos hasta que el usuario acciona un control.

46.2.3 Teoría de la transducción por capacitancia

En el reproductor de discos de vídeo de RCA se aplica la tecnología de la transducción por capacitancia para detectar la señales de vídeo y audio impresas en los surcos de un disco electrónico de capacitancia (CED). El CED es parecido a un disco de audio común en que la información de señales se coloca en surcos. Sin embargo, existen algunas diferencias importantes. Una es la densidad de surcos, como se ilustra en la figura 46-27. En un disco de audio, los surcos se espacian

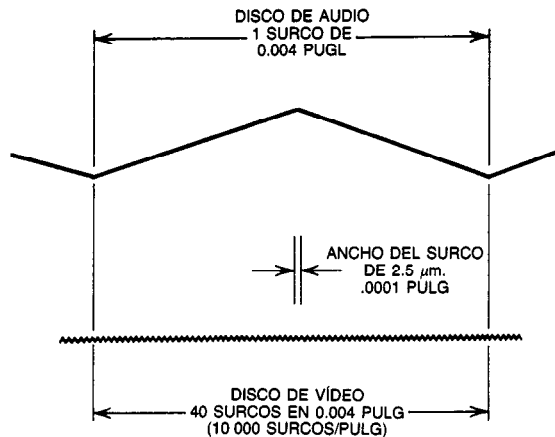


Fig. 46-27. Comparación de la densidad de surcos en discos de vídeo y de audio.

aproximadamente 0.004 pulg.; en el disco de vídeo los surcos están mucho más próximos. Se pueden colocar hasta 40 surcos en el espacio que existe entre dos surcos de un disco de audio. Esto quiere decir que se colocan aproximadamente 10 000 surcos en un radio de 1 pulg del disco de vídeo.

Las ondas portadoras de vídeo y audio moduladas en frecuencia se colocan en el disco variando la profundidad del surco con las señales portadoras, como se ilustra en la figura 46-28. El surco es muy pequeño, ya que sólo mide 2.5 μm de ancho. Por tanto, la modulación de la profundidad de la onda es necesariamente pequeña, más o menos de 850 Å de cresta a cresta como máximo para el vídeo.

En un sistema de discos de audio típico, la señal se detecta por el movimiento de un imán en una bobina pequeña, que es modulada por una aguja que recorre los surcos del disco. Esto genera una señal que varía en frecuencia y amplitud, con lo que se recupera la señal de audio original.

Debido al pequeño tamaño de los surcos del disco de vídeo y de la modulación de la profundidad, la detección de variaciones de amplitud y frecuencia por este método sería imposible. Por ello se creó el sistema de transducción por capacitancia.

En el sistema CED (disco electrónico de capacitancia), las señales de audio y vídeo se colocan en el disco a través de portadoras de FM. Esto elimina la necesidad de una detección exacta de variaciones de amplitud, ya que la amplitud de la señal de vídeo está representada ahora por la desviación de frecuencia de la señal portadora de FM. Por tanto, ahora sólo es necesario detectar la frecuencia de la señal en el disco de vídeo. Hacer esto por medios mecánicos, como en los discos de audio, seguiría siendo casi imposible, ya que la frecuencia de las señales en el disco de vídeo puede ser hasta de 7 MHz, lo que produce longitudes de onda tan cortas como 0.5 μm.

El sistema de detección de la capacitancia resuelve este problema. La longitud de la aguja captora o de transducción del disco de vídeo es varias veces mayor que la más larga longitud de onda grabada en el disco. Por tanto, cuando la aguja recorre la modulación en el surco, su posición vertical se mantiene constante. Se coloca un electrodo metalizado delgado en la superficie de salida (posterior) de la aguja. Este electrodo actúa como una placa del "capacitor". El disco de vídeo, que está hecho de un plástico conductor con una cubierta lubricante muy fina, actúa como la otra placa del capacitor. A medida que gira el disco, la distancia entre el borde inferior del electrodo de la aguja y la modulación del surco varía en función de la modulación. Esto hace cambiar la distancia entre las placas del capacitor a la frecuencia de la modulación, con lo cual cambia la capacitancia entre el electrodo de la aguja y el disco.

La capacitancia cambiante de aguja a disco a su vez modula una señal de UHF (915 MHz) en el conjunto resonador del brazo de captación. La señal de UHF modulada en amplitud resultante se detecta a nivel de "cresta", lo que genera una señal de salida que es una réplica en voltaje de las señales portadoras de audio y vídeo de FM grabadas en el disco. Estas portadoras de FM pueden desmodularse entonces para recuperar las señales de vídeo y audio.

Un valor típico de la fuerza vertical mínima de la aguja contra el disco (o simplemente fuerza vertical) de un sistema de audio, generada por el peso del cartucho fonocaptor y el brazo de tono, es aproximadamente de 0.5 a 2.5 g. En el sistema de discos de vídeo CED, como se ilustra en la figura 46-29, dicha fuerza es sólo

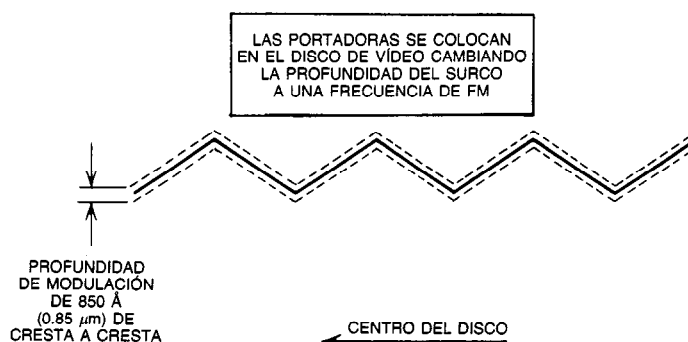


Fig. 46-28. Modulación del disco de vídeo.

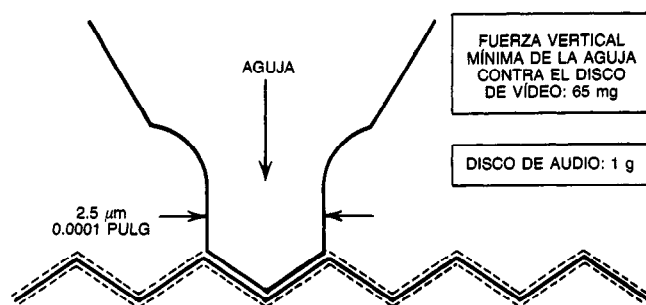


Fig. 46-29. Fuerza vertical mínima de la aguja contra el disco de vídeo.

de 65 mg (0.065 g). La fuerza vertical de la aguja del disco de vídeo es generada por el finísimo contacto móvil de la aguja y el pequeño peso de esta última. El contacto móvil conecta asimismo el electrodo de la aguja a la electrónica del resonador en el conjunto del brazo de captación.

46.2.4 Señales del disco de vídeo

A diferencia de los discos de audio, en los cuales las señales se graban directamente en el disco, las señales de vídeo y audio del disco de vídeo se utilizan para modular en frecuencia dos señales portadoras que, a su vez, se graban después en el disco. Esto reduce el intervalo dinámico de las señales que se van a grabar y también hace posible la colocación de más de una señal en el disco de vídeo mediante el empleo de dos o más frecuencias portadoras diferentes.

La señal de vídeo que se graba en el disco es una portadora de vídeo de 5 MHz modulada en frecuencia. El nivel del negro de la señal de vídeo provoca una desviación cero de la portadora o bien una frecuencia de 5 MHz. Las crestas de sincronización hacen que la frecuencia de la portadora de vídeo se desvíe a 4.3 MHz. La cresta del blanco de la señal de vídeo hace que la señal portadora de vídeo se desvíe a 6.3 MHz. Las bandas laterales que se generan a partir de la modulación en frecuencia de la portadora de 5 MHz se extienden desde 2 hasta 9.3 MHz.

La información de audio se debe colocar en el disco de vídeo al mismo tiempo que la información de vídeo. Las señales de audio se colocan en una frecuencia de portadora diferente de las señales de vídeo. Un canal de información de audio se coloca en una frecuencia de portadora modulada en frecuencia de 716 kHz. La señal de audio genera una desviación de frecuencia de ± 50 kHz. Por tanto, el espectro del disco de vídeo contiene una portadora de audio a 716 kHz y una portadora de FM de vídeo de 4.3 a 6.3 MHz con bandas laterales de 2 a 9.3 MHz (fig. 46-30).

Antes de la modulación de la portadora de vídeo de 5 MHz, la subportadora de crominancia de 3.58 MHz y las bandas laterales resultantes se convierten de manera descendente a 1.535625 MHz (1.53 MHz). Esto tiene varios efectos positivos. Primero, la conversión descendente de la información de croma permite el uso de una portadora de vídeo de frecuencia relativamente

baja (5 MHz). Las bandas laterales normales de una subportadora de 3.58 MHz excederían de los 4 MHz, con lo cual se incrementaría el nivel de ruido inherente del sistema. Segundo, la crominancia convertida en sentido descendente permite que la longitud de onda más corta grabada en el disco de vídeo sea comparativamente larga. Esto facilita la grabación de una hora de material de programa en cada cara del disco.

La crominancia convertida en sentido descendente se genera heterodinando la crominancia de 3.58 MHz con una señal de oscilador de 5.115170 MHz. La subportadora de crominancia resultante de 1.535625 MHz (1.53 MHz) se limita entonces en banda lateral a ± 500 kHz. Después se agrega información de luminancia a la crominancia convertida en sentido descendente para generar una señal de vídeo compuesta, con las señales de crominancia totalmente dentro del ancho de banda de luminancia. La conversión descendente de la información de croma a la mitad del espectro de luminancia se conoce algunas veces como sistema de subportadora oculta o de croma oculta.

La señal de vídeo compuesta resultante modula la portadora de vídeo de 5 MHz, generando la señal portadora de FM de vídeo de 4.3 a 6.3 MHz. Las bandas laterales de la portadora de vídeo se extienden de 2 a 9.3 MHz. La portadora de audio modulada en frecuencia de 716 ± 50 kHz se suma entonces a la portadora de FM de vídeo y se aplica a la cabeza cortadora de la máquina empleada para hacer negativos de discos de vídeo. La elaboración de un negativo de disco de vídeo es muy similar a la de un negativo o galvano de disco de audio, mediante el uso de un montaje mecánico para "cortar" la modulación en los surcos del disco. En la figura 46-31 se ilustra un diagrama de bloques global del proceso de grabación de un disco de vídeo.

46.2.5 Operación de la cubierta portadora

Debido al tamaño extremadamente pequeño de los surcos del disco de vídeo, la acumulación de polvo, granos de arena y otros agentes en la superficie del disco podría producir ruido indeseable en las señales de vídeo y audio detectadas. Para impedir que esto ocurra, cuando no está en uso el disco se encuentra contenido en una cubierta plástica conocida como contenedor. Cuando se va a reproducir el disco de vídeo, se inserta el contenedor en el reproductor, que retira el

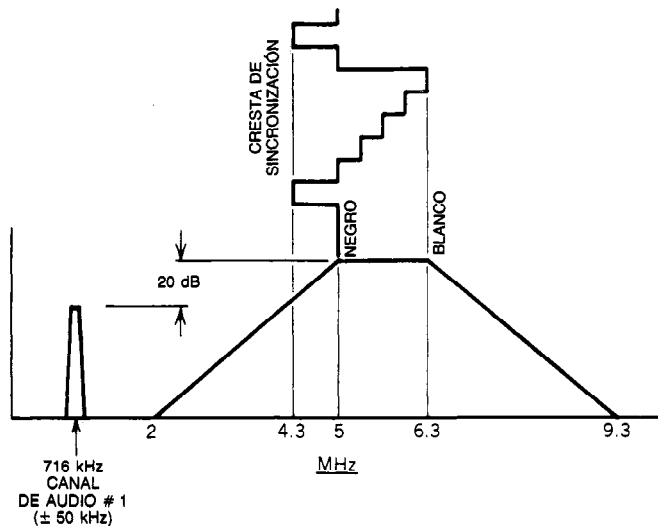


Fig. 46-30. Espectro de frecuencias del disco de vídeo (señal grabada en el disco).

conjunto del disco y eje del interior del contenedor. Entonces se extrae el contenedor vacío del reproductor para ejecutar el disco. Una vez que éste se ha reproducido se vuelve a insertar el contenedor vacío en el reproductor, y se extrae con el conjunto de disco y eje. De esta forma se impide la exposición directa del disco a la atmósfera exterior.

El contenedor evita asimismo que el usuario toque el disco, ya que las grasas y ácidos del sudor de los dedos podrían degradar la superficie lubricante de aquél, con una posible pérdida de señal.

Además de proteger el disco de vídeo de la contaminación y malos tratos, el contenedor sirve como un medio práctico de almacenamiento. Un empaque de tela suave en la abertura del contenedor forma un sello hermético y resistente al polvo que evita la entrada de contaminantes.

46.2.6 Descripción general del sistema

La mayor parte de los circuitos electrónicos del reproductor de discos de vídeo de RCA se pueden clasificar en dos categorías básicas: los circuitos de control del sistema y los de procesamiento de señales. Una microcomputadora es la parte principal de la electrónica de control del sistema. La computadora recibe entradas provenientes de los interruptores de funciones del usuario y, a su vez, controla la operación del reproductor. La microcomputadora de control del sistema es responsable asimismo de decodificar el código de información auxiliar digital (DAXI) que se suministra desde el disco de vídeo para producir la indicación del tiempo. La mayor parte de los componentes electrónicos de control del sistema están ubicados en el tablero PW 500, que está montado sobre la tornamesa del disco de vídeo.

Los circuitos de procesamiento de señales, contenidos principalmente en el tablero PW 3000 (y algunos

de ellos en el conjunto del brazo), son responsables de detectar la información de vídeo impresa en el disco, desmodular esta información, procesarla a través de un circuito filtro de característica en peine y modularla sobre los canales 3 o 4. La señal RF de televisión modulada se puede aplicar entonces a un receptor de televisión NTSC estándar. El sistema de procesamiento de señales completo está formado por seis circuitos integrados, que realizan la mayor parte de las funciones de procesamiento de señales. En la figura 46-32 se ilustra el diagrama de bloques operacional de un sistema de discos de vídeo "monoauricular" ("monaural").

Todos los componentes electrónicos que se necesitan para el procesamiento de la estereofonía en modelos aplicables se localizan en el tablero PW 4000. Por su parte, el tablero PW 5900 contiene toda la electrónica necesaria para dar cabida a las características de control remoto y "tacto suave" que se encuentran en ciertos modelos.

Operación funcional

La operación del reproductor de discos de vídeo es controlada en su totalidad por los circuitos de control del sistema. Las funciones operadas por el usuario de reproducción, avance y retroceso de acceso rápido, avance y retroceso de búsqueda visual, pausa y carga se introducen en la microcomputadora a través de los interruptores de funciones del usuario. La microcomputadora decodifica estas instrucciones de entrada y, a su vez, controla la electrónica para realizar esas funciones. El estado de todos los circuitos de procesamiento de señales es controlado por la salida "sin supresión" de la microcomputadora. Cuando la línea de no supresión pasa a un estado lógico "lo" (bajo), se desactivan todos los circuitos electrónicos.

La microcomputadora de control del sistema tiene asimismo control directo sobre el conjunto del brazo de captación. Este control implica la operación de un

servo, que hace avanzar el brazo durante una reproducción normal; la operación del levantador de la aguja, que eleva y hace descender la aguja cuando se inician diversas funciones; y los circuitos de retracción de la aguja, que hacen posible que el sistema ofrezca las características de búsqueda visual en modo de avance y retroceso. La microcomputadora controla también la dirección del servosistema. En los modos de operación de retroceso de acceso rápido y de retroceso de búsqueda visual, la microcomputadora instruye al servosistema para que opere en el modo de retroceso en vez de en el modo de operación normal (de avance).

La microcomputadora de control del sistema es asimismo responsable de la visualización del tiempo de reproducción. La información necesaria para ello se genera a partir de una señal DAXI que se graba en el disco de vídeo durante la fabricación. El código DAXI, que figura en la línea 17 de cada campo vertical, contiene un número de identificación de campo que es descifrado por la microcomputadora de control del sistema a fin de visualizar el tiempo transcurrido del programa en minutos.

En los modos de operación de avance y retroceso de acceso rápido, el código DAXI no está disponible debido a que la aguja ha sido levantada del disco. En este modo de operación, el visualizador del tiempo debe mantenerse de manera que el tiempo transcurrido aproximado del programa se pueda rastrear mientras el brazo se mueve en una u otra dirección a través del disco. Esto lo realiza un circuito "fotointerruptor", que verifica el tiempo transcurrido aproximado mediante el seguimiento de la posición relativa del brazo con respecto al disco.

Los componentes electrónicos situados en el conjunto del brazo de captación detectan la información de

vídeo contenida en el disco. El brazo contiene asimismo los componentes electrónicos que se encargan de proporcionar las características de avance y retroceso de búsqueda visual y también protección contra surcos cerrados. Estas características se ponen en funcionamiento mediante la activación de las bobinas "de retracción de la aguja", que cuando se activan hacen que la aguja salte dos surcos del disco de vídeo. También en el conjunto del brazo de captación se encuentra el transductor "extensor del brazo", que corrige las variaciones de la base de tiempo en las señales recuperadas de crominancia y luminancia debidas a combadura, excentricidad, cambios en la velocidad rotacional de la tornamesa o alguna combinación de estos factores.

La función principal de la electrónica de captación o transducción consiste en detectar las señales de vídeo colocadas en el disco. Esto se logra modulando un circuito resonador de UHF de 910 MHz con la capacitancia variante de la superficie del disco de vídeo. La variación en la capacitancia de la superficie del disco hace que se module la frecuencia central del resonador de 910 MHz, con lo cual la amplitud modula una señal fija de 915 MHz que atraviesa el resonador. Después se detecta esta señal a nivel de cresta, donde la señal resultante representa las variaciones de capacitancia en el disco de vídeo. Esta señal contiene las señales portadoras moduladas en frecuencia de vídeo y audio.

Las señales portadoras de vídeo y audio procedentes de los circuitos electrónicos del brazo se aplican a dos fases desmoduladoras de FM (una para sonido y otra para vídeo). El desmodulador de sonido decodifica la información de la portadora de audio y genera una señal de audio discreta, que después modula en frecuencia una portadora de 4.5 MHz en la etapa moduladora de RF.

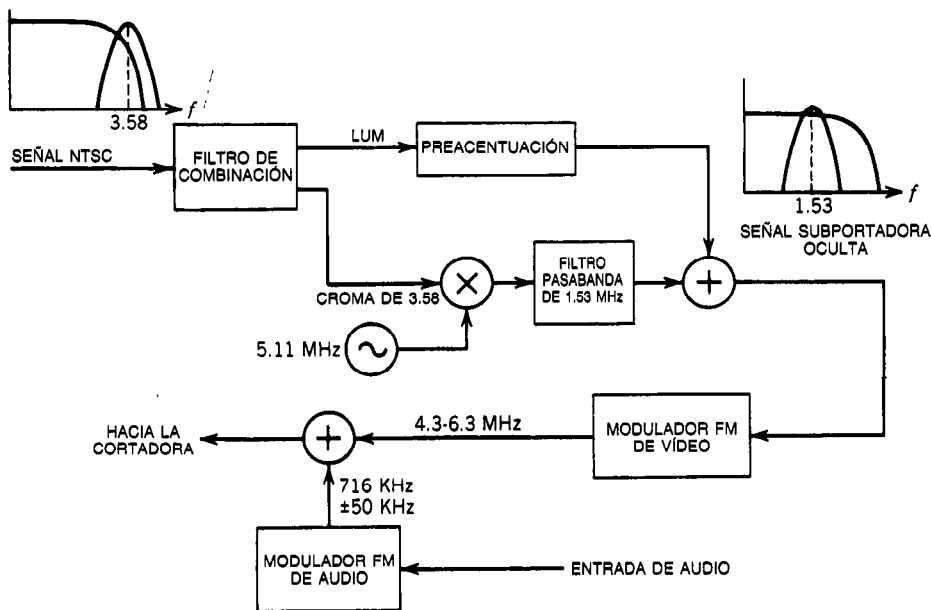


Fig. 46-31. Diagrama simplificado de bloques de la grabación de discos de vídeo.

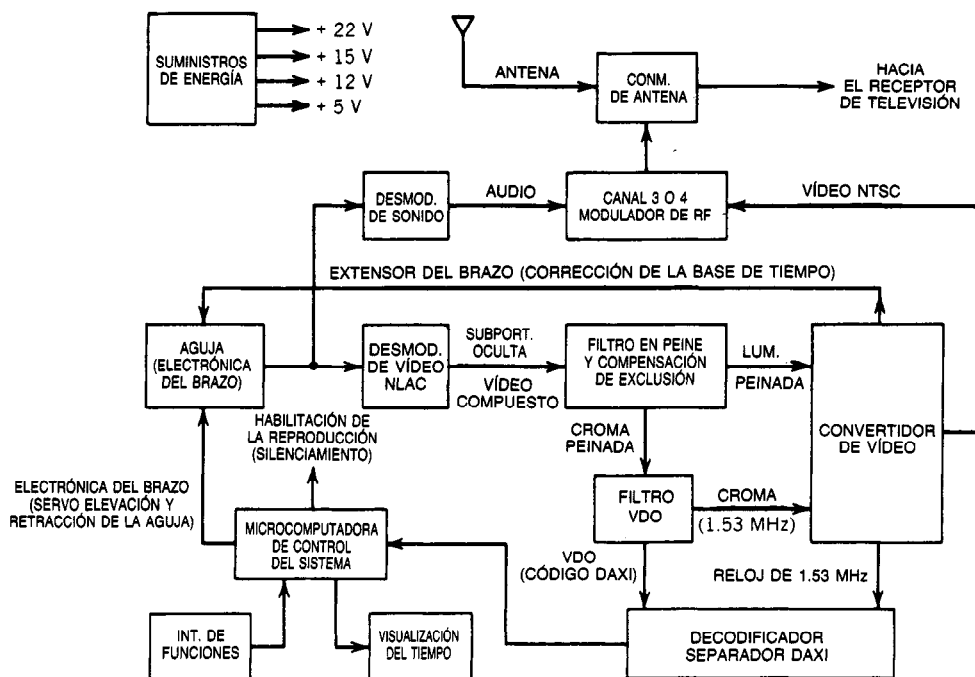


Fig. 46-32. Diagrama de bloques del reproductor de discos de vídeo de RCA. DESMOD. = desmodulador; CONM. = conmutador; SUBPORT. = subportadora; LUM. = luminancia; INT. = interruptores.

El desmodulador de sonido contiene asimismo un circuito corrector de defectos que reduce el ruido indeseable en audio si la portadora de sonido se pierde en forma momentánea debido a detritos microscópicos situados en la superficie del disco.

En el caso de un reproductor "estéreo" (SGT200/250) las señales portadora de vídeo y audio (dos), procedentes de los circuitos electrónicos del brazo, se aplican a tres fases desmoduladoras de FM. Una de las dos portadoras de audio, la de 716 kHz, se desmodula en el tablero PW 3000, tal como sucede en el reproductor de discos de vídeo monoauricular ("monaural"). La segunda portadora (o la de 905 kHz, o ambas) se desmodula y procesa en el tablero de procesamiento estéreo PW 4000. Una vez que se han recuperado ambos canales de audio (en el caso de un disco estereofónico), se aplican a un decodificador de reducción de ruido que está en el tablero estéreo.

Ambos desmoduladores de audio generan pulsos de defecto que se aplican a circuitos correctores de defectos (circuito de muestreo y retención), los cuales reducen el ruido indeseable en el audio si la portadora de sonido se pierde momentáneamente por la presencia de detritos microscópicos en la superficie del disco.

Antes de la desmodulación, la señal portadora de FM de vídeo se hace pasar por un circuito de corrección de apertura no lineal (NLAC). El circuito NLAC elimina pulsaciones de sonido de 716 kHz (de 905 kHz en las versiones "estereofónicas") en la reproducción debidas a la modulación en fase de la portadora de sonido contenida en la información de la portadora

de vídeo recuperada. El circuito NLAC elimina la modulación de la portadora de sonido invirtiendo la fase de la modulación y volviéndola a sumar a la señal original, con lo que se cancela la componente de modulación de 716 kHz (o 905 kHz) de la información de la portadora de vídeo.

La portadora de FM de vídeo se aplica entonces al desmodulador de vídeo, que detecta la portadora correspondiente. El desmodulador de vídeo contiene asimismo un circuito corrector de defectos o compensador de exclusión (o de caída de amplitud), que hace posible que se inserte la línea horizontal anterior cuando ocurra un defecto originado por la pérdida de la portadora.

La salida del desmodulador de vídeo, que es una señal de vídeo compuesta con subportadora de croma oculta, se aplica a un circuito filtro de característica en peine. Este filtro separa dinámicamente la información de crominancia y luminancia de la señal de vídeo compuesta. Se necesita un filtrado eficiente de color y luminancia, ya que la información de la subportadora de croma está oculta dentro del ancho de banda de la frecuencia de luminancia.

La salida del filtro de característica en peine es la crominancia y luminancia "peinadas". La señal de salida de crominancia peinada contiene asimismo información de luminancia de baja frecuencia y también la señal DAXI, que se transmite con cada campo vertical. El paso de banda recupera la señal de croma de 1 a 2 MHz; las señales restantes son separadas por filtros pasabajas. La información de luminancia de baja fre-

cuencia se recombina con la información de luminancia peinada para obtener la salida de luminancia. La señal DAXI se acopla a través de un decodificador/separador DAXI con la microcomputadora de control del sistema.

Una vez procesada por los circuitos filtros de característica en peine, la información de luminancia y crominancia se acopla a los circuitos convertidores de vídeo. El convertidor de vídeo convierte en sentido ascendente la información de crominancia de 1.53 MHz de nuevo a 3.58 MHz. A continuación, se combinan la información de croma de 3.58 MHz y la de luminancia. La señal de vídeo compuesta se aplica al modulador RF, donde se suma la portadora de audio FM y se genera una señal RF en el canal 3 o 4 para enviarse como salida a un receptor de televisión NTSC estándar.

También en la etapa de conversión de vídeo se genera la señal de impulso para los circuitos correctores de la base de tiempo del extensor del brazo. La señal de corrección se produce comparando la información de croma de 3.58 MHz que se genera en el convertidor de vídeo con una referencia de 3.58 MHz. Cualquier diferencia de fase o frecuencia entre las dos señales genera una señal de error, que se aplica a los circuitos del extensor del brazo. Estos circuitos operan un solenoide que mueve la aguja para mantener una velocidad de disco a aguja constante. La salida de los circuitos del extensor del brazo también se acopla con el oscilador convertidor para mantener la sincronización de fase entre la señal de color de 3.58 MHz convertida en sentido ascendente y el oscilador de referencia de 3.58 MHz.

Otro tablero que se encuentra en los reproductores de discos de vídeo más recientes es el tablero corrector de interferencia de pulsos (PIC). El objeto de este circuito es evitar que pulsos de radar y otros pulsos de RF fuertes en el intervalo de los 900 MHz interfieran en la operación del reproductor de discos de vídeo. El

circuito PIC detecta la presencia de estos pulsos e instruye al corrector de defectos en IC 3301 para que sustituya la línea anterior de información de vídeo.

Fuentes de alimentación CA y CD

Todos los circuitos electrónicos del reproductor de discos de vídeo de RCA están aislados de la línea de alimentación; es decir, están conectados "en frío" a tierra. Como se observa en la figura 46-33, se aplica una entrada de CA al tablero PW ACIN, en el cual un fusible de 1 A proporciona protección inicial. El interruptor de energía CA, S2, es controlado por la palanca de funciones. Se aplica energía CA al transformador de energía CA T1 cuando el interruptor de función está en la posición de carga o reproducción. El transformador de energía T1 es protegido asimismo por F2, un fusible de 1/4 A.

El secundario de T1 contiene dos devanados, uno que genera 9 V rms, utilizado para producir un suministro de 5 V para el sistema de control de la microcomputadora, y otro que genera 18 V rms y suministra 22 V de CD.

También se aplica energía CA de S2 al circuito de reproducción de CA, S4. El interruptor de reproducción de CA es controlado por la palanca de funciones, y sólo se cierra en la posición de reproducción. El interruptor de reproducción de CA se abre en la posición de carga.

El interruptor de reproducción de CA S4 se conecta al interruptor de detección del eje de CA S8, que está en serie con el motor de la tornamesa. El interruptor de detección del eje de CA se cierra solamente cuando se ha insertado (y una vez se ha retirado vacía) una funda portadora en el reproductor, dejando en él el eje y el disco. El cierre del interruptor de energía, el interruptor de reproducción de CA y el interruptor de detección del eje de CA activan entonces el motor de la tornamesa.

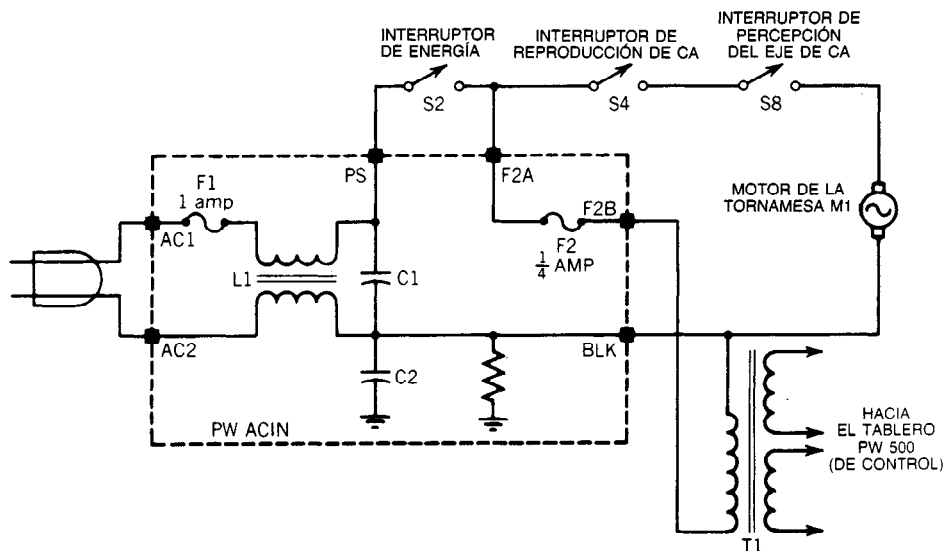


Fig. 46-33. Entrada de CA y energía de la tornamesa.

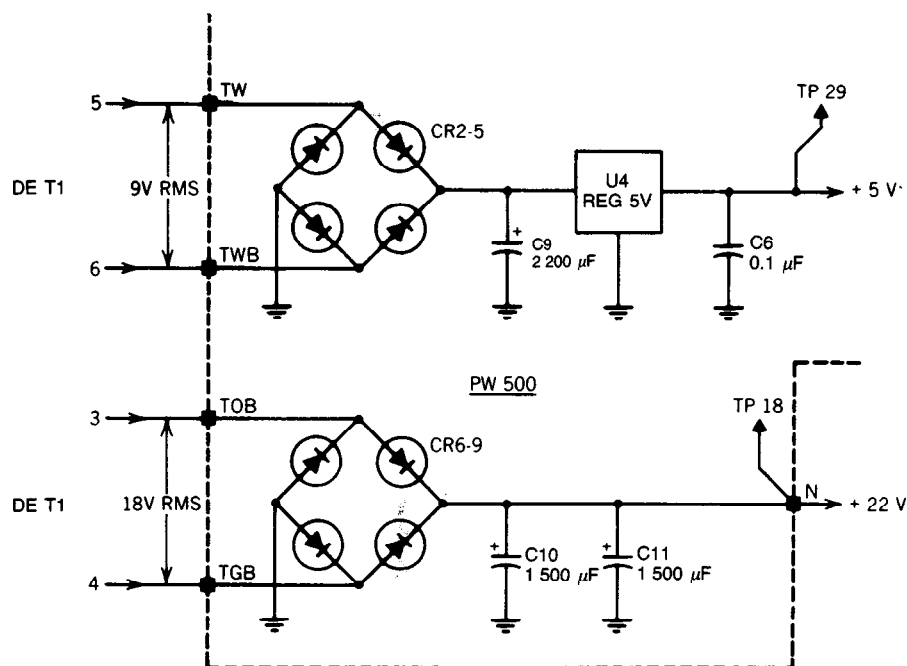


Fig. 46-34. Generación de suministros de +5 y +22 V B +.

El motor de la tornamesa es un motor de CA bipolar de polo sombreado que impulsa la tornamesa a 450 rpm. La sincronización entre la tornamesa y la línea de alimentación es conservada por dos polos magnéticos conectados al motor, que impulsa un anillo magnético de 16 polos situado en el interior de la base de la tornamesa. Los polos magnéticos del motor que impulsa el anillo magnético en torno a la base de la tornamesa garantizan que ésta mantenga una relación de fase constante con la entrada de CA de 60 Hz. Esto da por resultado una frecuencia de campos de vídeo muy estable de 60 Hz.

Como se ilustra en la figura 46-34, los 9 V rms (eficaces) que se generan entre las terminales 5 y 6 de T1 se aplican a los diodos del rectificador en puente CR 2, CR 3, CR 4 y CR 5 ubicados en el tablero PW 500. El rectificador en puente proporciona un voltaje de CD que es filtrado por el capacitor C9, un capacitor electrolítico de 2 200 μF . La CD filtrada se aplica entonces a un regulador de 5 V de tres terminales (U4) que genera un voltaje de salida de 5 V, el cual es filtrado por C6. El suministro de 5 V se utiliza para alimentar la microcomputadora de control del sistema, la pantalla LED y otros circuitos de control del sistema.

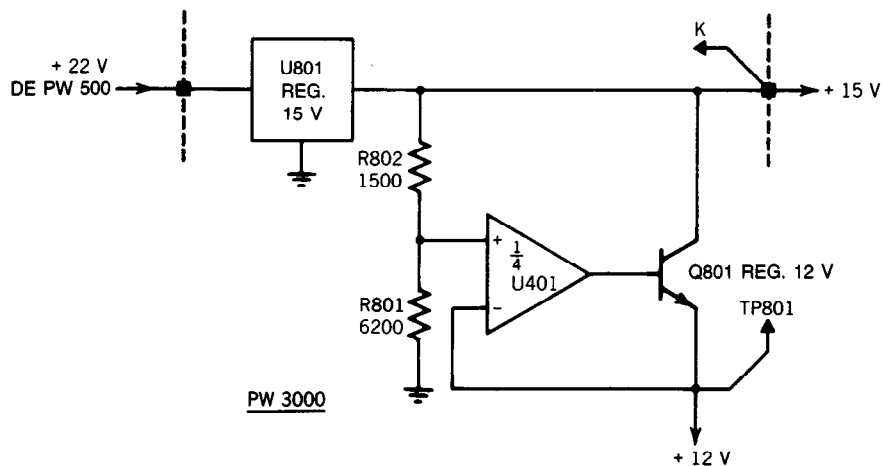


Fig. 46-35. Generación de suministros de +12 y +15 V.

Se aplican 18 V de CA de las terminales 3 y 4 del transformador T1 a los diodos del rectificador en puente CR 6, CR 7, CR 8 y CR 9 en el tablero PW 500. La salida de ese rectificador en puente se aplica a los capacitores C10 y C11 (dos capacitores electrolíticos de 1 500 μ F en paralelo). El resultado es un suministro de 22 V B + bien filtrado, que utilizan los circuitos servo y de accionamiento de retracción de la aguja.

Los 22 V de CD generados en el tablero PW 500 se suministran asimismo al tablero PW 3000 (terminal N), donde se aplican a la terminal de entrada del U801, un regulador de tres terminales de 15 V (fig. 46-35). La salida regulada de 15 V del U801 se puede medir en la terminal K. Por medio de U401 y Q801 se genera un suministro regulado de 12 V a partir del suministro de 15 V. Una red divisora que consta de R801 y R802 proporcionan aproximadamente 12 V en la entrada positiva de un amplificador operacional (amp op). La salida del amp op activa la base del transistor regulador Q801 de 12 V. La salida regulada de 12 V se vuelve a acoplar después a la entrada negativa del amp op, manteniendo una salida constante de 12 V en TP 801. El amp op es parte del U401, que es un paquete de cuatro amplificadores operacionales.

La mayor parte de los componentes eléctricos del reproductor de discos de vídeo están situados en uno de varios subconjuntos grandes. El tablero PW 3000 contiene todos los circuitos de procesamiento de señales y también las etapas modulares de RF. Este tablero está unido a la parte inferior del aparato y se puede colocar sin dificultad en posición de servicio desmontando los tornillos de ensamble y doblando el tablero a un lado del reproductor. En este punto, todos los conjuntos son accesibles desde la parte superior del aparato.

El tablero PW 500, que contiene todos los componentes de control del sistema, está montado en la parte superior del aparato con los componentes boca abajo hacia la tornamesa. Este tablero se puede desmontar y colocar al lado del reproductor o invertirse e insertarse

en una ranura en la parte superior del aparato. Ambos fusibles, así como los circuitos de entrada de CA, están en el tablero PW ACIN, situado directamente debajo del lado derecho de la tornamesa.

El conjunto del brazo de captación contiene el circuito resonador de 915 MHz, el preamplificador y circuitos AFT, así como el extensor del brazo, circuito de retracción de la aguja y servocircuitos sensores. El cartucho de la aguja, reemplazable, está contenido asimismo en el montaje del brazo de captación. El cartucho se reemplaza fácilmente a través de una puerta situada encima del conjunto del brazo.

El servomotor y el conjunto de engranes de reducción están situados hacia la parte del fondo del reproductor. Una cremallera (parte del conjunto del brazo de captación) es accionada por el servomotor, que desplaza el brazo sobre el disco de vídeo. El conjunto de servo de engranes de reducción incorpora un embrague que es operado por la palanca de funciones. El embrague desconecta los engranes de accionamiento del servo para permitir que el brazo se desplace fácilmente durante la extracción (retirada) del disco.

LIBROS Y REVISTAS

- C. Bensinger, *The Home Video Handbook*, 3.^a ed., Howard Sams, Indianapolis.
- D. Chesire, *The Video Manual*, Van Nostrand, Nueva York.
- W. Clason, *Dictionary of Television and Video Recording*, Elsevier, Nueva York.
- W. Clason, *Home Video Yearbook*, Knowledge Industry, White Plains, NY.
- D. Ennes, *Television Broadcasting; Tape and Disc Recording Systems*, Howard Sams, Indianapolis.
- R. Robinson, *The Video Primer: Equipment Production and Concepts*, Putnam, Nueva York.
- E. Sigal y colaboradores, *Video Discs: The Technology, the Applications and the Future*, Van Nostrand, Nueva York.

DÉCIMA PARTE

Comunicaciones

CAPÍTULO 47

Conceptos sobre comunicaciones

Yeng S. Kuo
Kwei Tu

Lockheed Engineering and Management
Services Company, Inc.
Houston, Texas

Rodger E. Ziemer
University of Colorado-Colorado Springs
Colorado Springs, Colorado

William H. Tranter
University of Missouri-Rolla
Rolla, Missouri

47.1 Conceptos sobre comunicaciones modernas

- 47.1.1 Introducción
- 47.1.2 Modelo típico de un sistema de comunicación
- 47.1.3 Ruido y distorsión
- 47.1.4 Medidas de la calidad

47.2 Teoría del ruido

- 47.2.1 Fuentes de ruido
- 47.2.2 Parámetros y características

47.3 Teoría de la información

- 47.3.1 Conceptos básicos
 - 47.3.2 Codificación de la fuente
 - 47.3.3 Comunicaciones fiables con ruido
-

47.1 CONCEPTOS SOBRE COMUNICACIONES MODERNAS

Yeng S. Kuo

47.1.1 Introducción

El objetivo de las comunicaciones es transmitir información de un punto a otro. La era de la comunicación eléctrica dio comienzo cuando Samuel Morse transmitió el primer mensaje telegráfico por un alambre de 16 km de largo. Diez años después de que Hertz verificara la teoría de Maxwell en 1877, Marconi demostró satisfactoriamente y patentó un sistema telegráfico inalámbrico completo. Durante la Segunda Guerra Mundial se desarrollaron los sistemas de radar y microondas, junto con métodos electrónicos y estadísticos mejorados para analizar problemas de extracción de señales. Los cimientos de las comunicaciones modernas se colocaron sólidamente en este periodo.

Una característica inherente de la comunicación eléctrica es la presencia de incertidumbre, debida en parte a la presencia inevitable de perturbaciones indeseables en las señales, a las que suele llamarse ruido. El ruido ha sido un problema siempre presente desde los primeros días de la comunicación eléctrica. Durante mucho tiempo se pensó que el ruido imponía una restricción inevitable a la exactitud de las comunicaciones. En virtud de la naturaleza estocástica del ruido, Rice empleó procedimientos de análisis probabilístico para analizar las comunicaciones en presencia de ruido. Wiener y Kolmogoroff aplicaron métodos estadísticos a problemas de detección de señales.

En 1948 Shannon publicó su famoso artículo "*A Mathematical Theory of Communications*", en el cual reconocía la naturaleza estocástica de la información. Shannon planteó el siguiente problema: Dado un conjunto de posibles mensajes que una fuente puede producir de manera aleatoria, ¿cómo deben representarse o codificarse los mensajes para conducir de la mejor forma la información con un sistema dado? Shannon demostró que, para una fuente de información dada que transmite en un canal de comunicación también dado, existe una técnica de codificación tal que el error puede hacerse arbitrariamente pequeño, a pesar de la presencia de ruido, mientras la velocidad de transmisión de la información no exceda la capacidad del canal. Este trabajo marcó el inicio de las comunicaciones modernas. Aunque Shannon demostró que es posible la transmisión libre de error en un canal ruidoso empleando la codificación adecuada, su teoría no revelaba el tipo de técnica de codificación que debería emplearse. En la década de 1950 se desarrollaron numerosas teorías de la codificación; en este periodo el interés se concentró en el diseño de códigos, decodificación en bloque, estructura de peso y cotas a la distancia. En la década de 1960 Fano desarrolló la teoría de la codificación convolucional con decodificación sucesiva, y Viterbi la teoría de la decodificación de máxima probabilidad. En este tiempo también se hicieron análisis sobre el diseño más eficiente de señales y sobre técnicas de modulación y desmodulación. Con el rá-

pido desarrollo en la década siguiente de las tecnologías avanzadas de electrónica de estado sólido, tales como la pastilla de circuito integrado a muy grande escala (VLSI) de ultraalta velocidad y los poderosos microprocesadores, salieron a la luz muchas técnicas complejas de codificación/decodificación y modulación/desmodulación. Con ello se inició una nueva era de la comunicación eléctrica, cuyas aplicaciones potenciales sólo son limitadas por las necesidades; las aspiraciones y la imaginación de usuarios y diseñadores.

47.1.2 Modelo típico de un sistema de comunicación

En la figura 47-1 se presenta un modelo típico de sistema de comunicación. Como se ilustra, un **sistema de comunicación** se define como el proceso de transferir información de una fuente desde un punto a otro. Los elementos funcionales son: fuente de información y destino, transmisor y receptor, y el canal.

Fuente de información

La fuente de información produce un mensaje o una sucesión de mensajes por transmitir a la terminal receptora o destino. Existen muchos tipos de fuentes de información y, por tanto, los mensajes asumen diversas formas. Por lo general es posible clasificar ampliamente los mensajes en tres tipos:

1. Señales analógicas (onda de forma continua), que pueden modelarse como funciones de variables de tiempo continuo. El habla, la música, la televisión y las mediciones de temperatura son buenos ejemplos de señales analógicas.
2. Señales digitales, que consisten en símbolos discretos, como la salida de una computadora digital, la voz digitalizada y la televisión digitalizada.
3. Señales pulsátiles, que consisten en una sucesión de pulsos estrechos, tales como los empleados en las aplicaciones de radar y otras formas de detección.

El mensaje (señal) producido por una fuente no necesariamente debe estar en forma eléctrica, y puede requerirse un transductor para convertir el mensaje en una forma de transmisión más conveniente. Para los fines de esta sección se supone que los mensajes están en forma eléctrica.

Transmisor

El objetivo del transmisor es convertir el mensaje en una forma adecuada para su transmisión en el canal. Este proceso incluye una o más de las operaciones que se definen a continuación:

- **Filtrado.** Consiste en limitar el ancho de banda de la fuente de información o modificar la onda de forma.
- **Amplificación.** Consiste en incrementar la amplitud de la señal a un nivel apropiado para el procesamiento y transmisión.

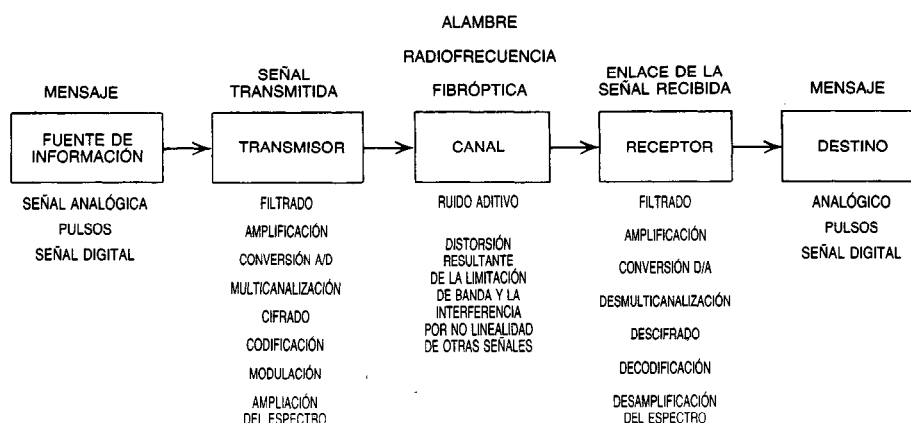


Fig. 47-1. Modelo de un sistema genérico de comunicación.

- **Modulación.** Con esta operación se produce una onda de forma adecuada para el canal.

Además de estas tres operaciones básicas, muchos transmisores también pueden realizar operaciones especiales, que se definen como sigue:

- **Conversión de analógico a digital.** Consiste en convertir una onda de forma continua (señal analógica) en una de forma digital.
- **Multicanalización.** Proceso de combinar dos o más señales de información separadas en una sola señal transmitida.
- **Cifrado.** Protección de datos contra alteración o intromisión.
- **Codificación.** Asignar códigos a los datos digitales para su detección y corrección.
- **Ampliación del espectro.** Expansión del espectro de radiofrecuencia de la señal transmitida con fines de controlar el acceso y reducir la densidad espectral (sección 48.5).

Canal

El canal es el medio de transmisión del sistema de comunicación. Suele consistir en un enlace de espacio libre (con antenas), un par de alambres, un cable o una fibra óptica. Un canal de comunicación tiene las siguientes características:

- **Atenuación.** Una función de la distancia recorrida por la señal (sección 48.2).
- **Distorsión.** Causada por desvanecimiento (debilitación de la señal, sección 48.7), interferencia, no linealidad o limitación de banda.
- **Ruido.** Generalmente modelado como una característica del canal, pero no necesariamente originada en él (sección 47.2).

Todas estas características del canal tienden a degradar la calidad de la señal que llega al receptor, e introducen la posibilidad de que la señal transmitida no se reciba con completa fidelidad.

Receptor

El receptor recupera del canal la señal, junto con el ruido, la distorsión y la interferencia acompañantes, y reconstruye con tanta exactitud como sea posible la señal de información original.

47.1.3 Ruido y distorsión

El ruido y la distorsión son las dos causas principales de degradación del funcionamiento en los sistemas de comunicación. **Ruido** es un nombre general dado a diversos fenómenos que se analizan en detalle en la sección 47.2. La presencia de ruido en los sistemas de comunicación es lo que más investigación ha motivado en la teoría de las comunicaciones, puesto que sin ruido sería posible crear un sistema de comunicación perfecto (libre de errores). Todas las teorías de la información, de la codificación y de la estimación y detección estadísticas tratan sobre el problema de extraer una señal deseada de un medio ruidoso.

La **distorsión** es una alteración de la señal debido a la respuesta imperfecta del sistema de comunicación. A continuación se enumeran los tipos más comunes de distorsión en tales sistemas:

- **Distorsión de amplitud.** Causada por la respuesta no uniforme de frecuencia en la banda de interés.
- **Distorsión de fase.** Debida a la respuesta no uniforme de fase en la banda de interés.
- **Distorsión armónica.** Por efecto de la respuesta no lineal.
- **Distorsión intermodulación.** Causada por una combinación no lineal de múltiples componentes de frecuencia en la señal.

Un tipo más de distorsión es la causada por la interferencia de otros sistemas de comunicación, ya sea inadvertida o intencional ("emparedamiento"). En los modernos sistemas de comunicación se han desarrollado técnicas para reducir los efectos de este tipo de distorsión, como se analiza en la sección 48.5.

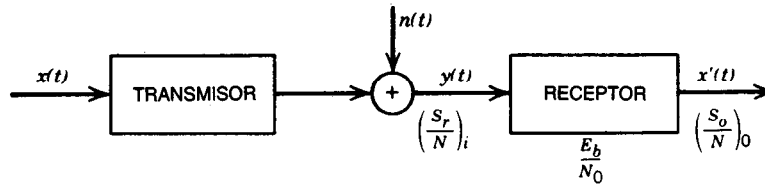


Fig. 47-2. Modelo analítico de un sistema de comunicación.

La mayor parte de las técnicas para contrarrestar los efectos de la distorsión y el ruido se basan en el principio de incrementar el ancho de banda de la señal transmitida y la complejidad del procesamiento de la señal en el receptor.

Con respecto a la figura 47-1, la práctica anterior significa que, si bien es deseable conservar tan pequeño como sea posible el ancho de banda del mensaje para reducir los efectos finales del ruido, es posible e incluso deseable ocupar un ancho de banda mucho mayor en el canal de comunicaciones para combatir los efectos de ruido, distorsión e interferencia.

47.1.4 Medidas de la calidad

Un sistema de comunicación ideal está libre de distorsión y transmite sin errores. En otras palabras, en el receptor puede recuperarse una réplica idéntica de la señal transmitida, posiblemente a escala y con retraso de tiempo.

Por desgracia el ruido es inevitable en el sistema, y al implementarlo suele presentarse la distorsión. De esta manera, es necesario examinar los efectos del ruido y la distorsión en el funcionamiento del sistema, y emplear algunas medidas cuantitativas para describir la calidad de la señal recuperada.

Señales analógicas

En la figura 47-2 se presenta de forma analítica el diagrama de bloques simplificado de un sistema de comunicación.

En el caso de un sistema analógico, el mensaje de entrada $x(t)$ es una onda de forma continua (analógica); $y(t)$ es la suma del mensaje recibido, atenuado por el paso a través del canal; $n(t)$ es el ruido del canal; y $x'(t)$ es el mensaje recuperado. Se supone que $x(t)$ está limitado en cuanto a su banda de transmisión y tiene una frecuencia superior máxima W , que el ruido es gaussiano (sección 47.2) y que el canal está libre de distorsión.

La razón de señal sobre ruido a la entrada del receptor puede representarse como:

$$(S/N)_i = \frac{\overline{y^2(t)}}{\overline{n^2(t)}} = \frac{S_r}{N_0 B}$$

donde S_r = potencia de la señal recibida

N_0 = densidad espectral de potencia del ruido

B = ancho de banda del receptor

y la barra superior indica un valor promedio.

La razón de señal sobre ruido del mensaje recuperado es entonces:

$$(S/N)_o = \frac{\overline{x'^2(t)}}{N_0 W} = \frac{S_o}{N_0 W}$$

donde S_o es la potencia de la señal a la salida del receptor.

El término $(S/N)_o$ es la medida de calidad usualmente aceptada de la señal recibida y puede expresarse como función de la razón de señal sobre ruido para la entrada, $(S/N)_i$. En un sistema de modulación lineal $(S/N)_o \leq (S/N)_i$, mientras que en uno no lineal, por ejemplo, uno de FM, el rendimiento puede mejorarse incrementando el ancho de banda de transmisión (sección 48.1).

En la figura 47-3 se presenta la comparación $(S/N)_o$ y $(S/N)_i$ para AM y FM con dos valores del índice de modulación β .

Algunas veces es posible reducir la distorsión lineal en un sistema de comunicación empleando ecualizadores. Estos dispositivos tienen una función de transferencia que al ser multiplicada por la propia del sistema produce una respuesta compuesta sin distorsión. Para aplicar tales técnicas es necesario conocer la función de transferencia real del sistema, o emplear algunas técnicas adaptativas para determinarla.

Por otro lado, la distorsión no lineal sólo puede remediarse en casos específicos, por ejemplo, cuando la no linealidad se introduce deliberadamente. Tal es el caso cuando se emplea un compansor (de compresor/expansor) para mejorar el intervalo dinámico de un sistema.

Las mediciones típicas en el caso de la distorsión no lineal son las siguientes:

$$\begin{aligned} \text{Distorsión armónica total} &= \\ &= \left(\frac{\text{potencia armónica total}}{\text{potencia fundamental total}} \right)^{1/2} \times 100\% \end{aligned}$$

$$\begin{aligned} \text{Distorsión de intermodulación total} &= \\ &= \left(\frac{\text{potencia de intermodulación total}}{\text{potencia fundamental total}} \right)^{1/2} \times 100\% \end{aligned}$$

Señales digitales

El objetivo de un sistema de comunicación digital es transmitir un mensaje discreto en un tiempo prescrito con un error mínimo (o mínima probabilidad de error). La medida de calidad aceptada en un sistema de comunicación digital es la probabilidad de error, ya sea de un mensaje, un símbolo o un solo dígito binario (bit). En el modelo analítico de la figura 47-2, el mensaje de entrada $x(t)$ es una sucesión de mensajes bi-

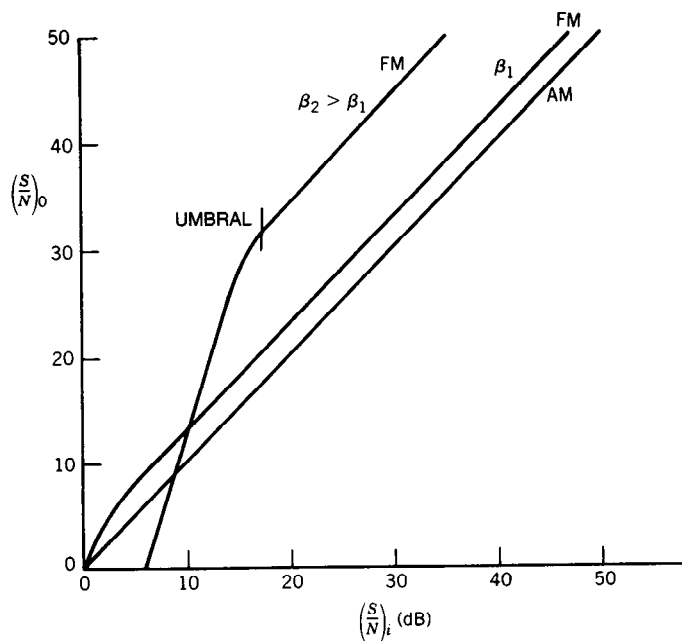


Fig. 47-3. Razón de señal sobre ruido de la salida de sistemas de FM y AM.

narios, y la razón de señal sobre ruido para la entrada se representa por medio de la razón de energía del símbolo (o bit) sobre densidad espectral del ruido, E_b/N_0 .

En la figura 47-4 se presenta la probabilidad media de error para la detección de un solo bit en presencia de ruido gaussiano blanco aditivo (que se explica en la sección 47.2). La curva mostrada es para un sistema óptimo de transmisión binaria. El rendimiento puede mejorarse (menor probabilidad de error para una razón E_b/N_0 dada) empleando símbolos no binarios y técnicas avanzadas, que se tratarán con detalle en la sección 48.5.

LIBROS Y REVISTAS

- S. O. Rice, "Mathematical Analysis of Random Noise", en N. Wax, ed., *Selected Papers on Noise and Stochastic Processes*, Dover, Nueva York.
- M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, Nueva York.
- M. Schwartz, W. R. Bennett y S. Stein, *Communication Systems and Techniques*, McGraw-Hill, Nueva York.
- H. Taub y D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, Nueva York.
- H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Parte 1, Wiley, Nueva York.
- A. Viterbi, *Principles of Coherent Communications*, McGraw-Hill, Nueva York.
- J. Wozencraft e I. Jacobs, *Principles of Communication Engineering*, Wiley, Nueva York.
- R. E. Ziemer y W. H. Tranter, *Principles of Communications*, Houghton Mifflin, Boston.

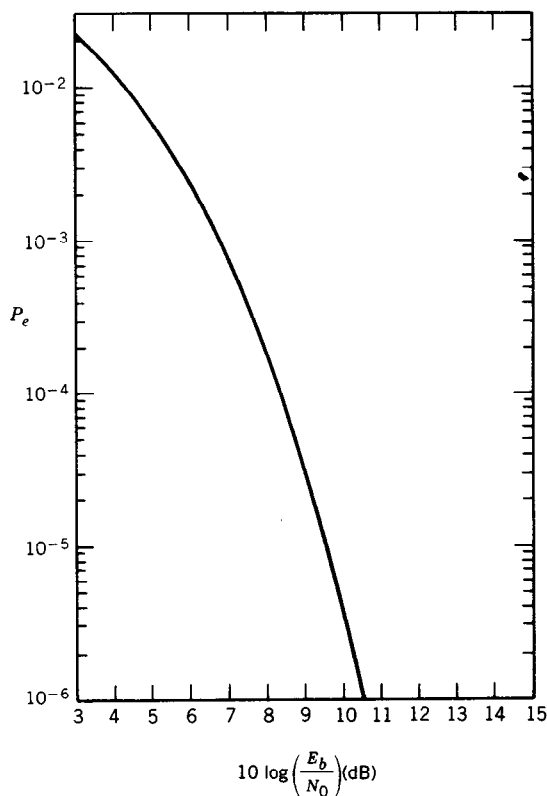


Fig. 47-4. Probabilidad de error de un sistema binario óptimo.

47.2 TEORÍA DEL RUIDO

Kwei Tu

47.2.1 Fuentes de ruido

El ruido y la interferencia causada por otros sistemas de comunicaciones son dos factores que limitan el rendimiento de todos los sistemas. El ruido proviene de diversas fuentes; las más importantes son:

1. Fuentes externas, como ruido atmosférico, ruido galáctico y ruido generado por el hombre.
2. Ruido térmico generado por pérdidas disipativas en el sistema de la línea de recepción a transmisión.
3. Fuentes internas del sistema receptor.

El nivel de ruido suele expresarse como temperatura en grados Kelvin (K) o decibeles (dB) respecto a una temperatura de ruido estándar de 290 K.

Fuentes externas

Ruido atmosférico. Este ruido es producido principalmente por rayos en tormentas eléctricas. De este modo, el nivel de ruido depende de la frecuencia, hora del día, tiempo meteorológico, estación del año y localización geográfica. Este ruido, sujeto a variaciones asociadas con las áreas de tormenta locales, por lo general disminuye al aumentar la latitud. Es particularmente intenso durante las estaciones lluviosas en regiones como el Caribe, Antillas Orientales, África ecuatorial y el norte de la India. Sin embargo, como se observa en la figura 47-5, este ruido atmosférico no es

una contribución importante al ruido del sistema en el caso de la frecuencia ultraalta (UHF) y frecuencias mayores. El ruido debido a la absorción atmosférica puede adquirir importancia por encima de 1 GHz cuando se emplean amplificadores de bajo ruido en el receptor. En la figura 47-6 se indica la temperatura del cielo para un haz infinitamente agudo. Dado un ángulo de elevación de 0.1 rad, la temperatura efectiva típica del cielo es de 13 K para 1 GHz y de 25 K para 10 GHz; para 20 GHz esa temperatura aumenta hasta 80 K.

Ruido galáctico. Éste puede definirse como el ruido de radiofrecuencia producido por perturbaciones originadas fuera de la Tierra o su atmósfera. Las causas principales de tal ruido son el Sol y un gran número de fuentes discretas de radiofrecuencia, distribuidas en su mayor parte en el plano galáctico. El ruido galáctico que llega a la superficie terrestre varía aproximadamente de 15 MHz a 100 GHz, y está limitado por la absorción ionosférica en el extremo inferior del espectro y por la absorción atmosférica en el superior. En la práctica, la importancia del ruido galáctico es restringida a frecuencias no menores de aproximadamente 18 MHz por el ruido atmosférico, y a frecuencias no mayores de 500 MHz por la temperatura de ruido del receptor y la ganancia de la antenna (fig. 47-5). No obstante, en el caso de una antenna receptora de alta ganancia dirigida hacia el Sol, la temperatura de ruido de la antenna puede superar los 290 K a frecuencias tan altas como 10 GHz. En la figura 47-7 se presenta el nivel de ruido galáctico en decibeles respecto a una temperatura de ruido de 290 K cuando la antenna es un dipolo de media onda. En los niveles de ruido que se presentan no se supone absorción atmosférica, y se consideran las siguientes fuentes de ruido galáctico:

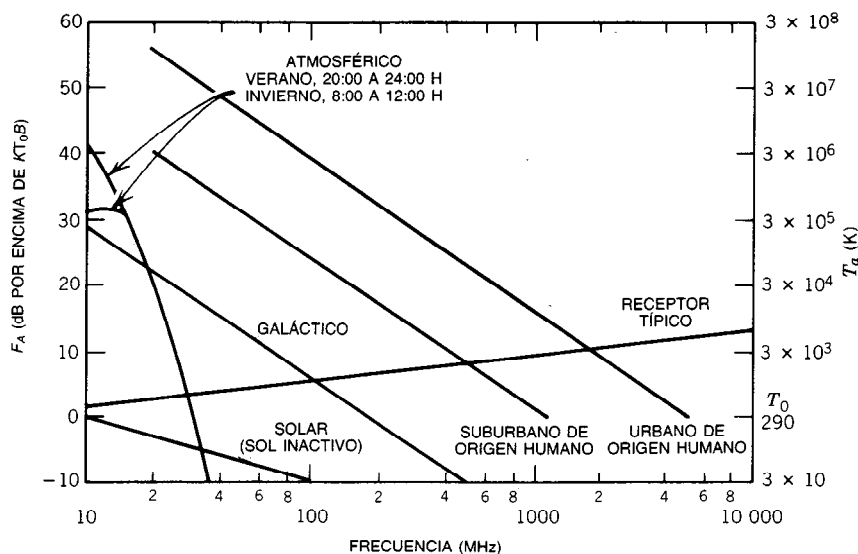


Fig. 47-5. Valores medianos de la potencia promedio de ruido esperada de distintas fuentes. (Tomado de CCIR 10.^a Plenary Assembly, "World Distribution and Characteristics of Atmospheric Radio Waves", Informe núm. 322, Ginebra.)

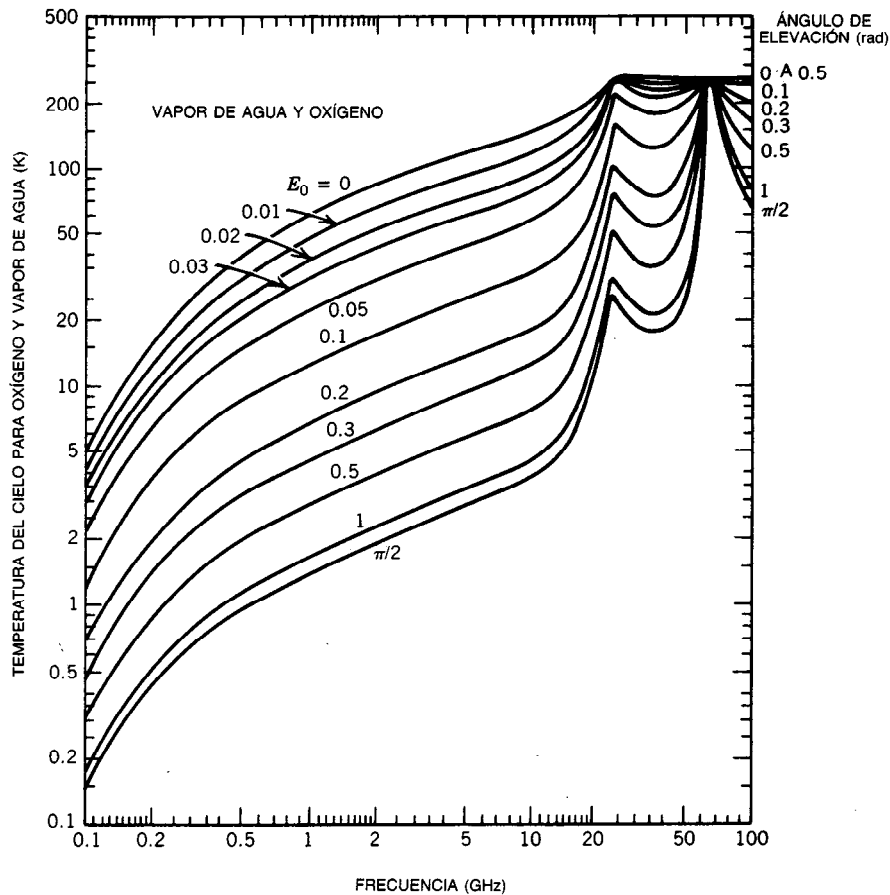


Fig. 47-6. Temperatura de ruido celeste debida a la rerradiación por el oxígeno y vapor de agua. (Tomado de CCIR 10.^a Plenary Assembly, "World Distribution and Characteristics of Atmospheric Radio Waves". Informe núm. 322. Ginebra.)

- **Plano galáctico.** Ruido del plano galáctico en la dirección del centro de la galaxia. Los niveles de ruido procedente de otras partes del plano de la galaxia pueden ser hasta 20 dB menores que los dados en la figura 47-7.
- **Sol inactivo o en reposo.** Ruido procedente del Sol "en reposo"; esto es, el ruido solar en tiempos de poca o nula actividad de las manchas solares.
- **Sol activo.** Ruido procedente del sol "activo", aludiendo a los periodos de actividad de las manchas y la corona solares.
- **Casiopea.** Ruido de una fuente discreta de alta intensidad de ruido cósmico, conocida como Casiopea. Ésta es una de las más de cien fuentes discretas conocidas, cada una de las cuales subtendiendo en la superficie terrestre un ángulo de menos de medio grado.

Los niveles de ruido cósmico recibidos por una antena dirigida a una fuente de ruido pueden estimarse corrigiendo los niveles de ruido relativos con un dipolo

de media onda (fig. 47-7) para la ganancia de antena receptora obtenida en la fuente de ruido. En virtud de que el plano galáctico es una fuente extendida de ruido no uniforme, no es posible obtener ganancias de antena de espacio libre, sino que el límite es aproximadamente de unos 10 a 15 dB. Sin embargo, en el Sol y otras fuentes de ruido galáctico son posibles ganancias de antena de 50 dB o más.

Ruido producido por el hombre. La amplitud del ruido producido por el hombre disminuye al aumentar la frecuencia, y varía considerablemente con el lugar. Es debido en su mayor parte a motores eléctricos, anuncios de neón, líneas eléctricas y sistemas de ignición localizados en algunos cientos de metros de la antena receptora; sin embargo, ciertos aparatos médicos de alta frecuencia y líneas de transmisión de alto voltaje pueden causar interferencias a distancias mucho mayores. El nivel promedio de la potencia del ruido generado por el hombre puede ser mayor en 16 dB o más en áreas urbanas que en suburbanas; en localidades rurales remotas el nivel puede ser 15 dB menor que el

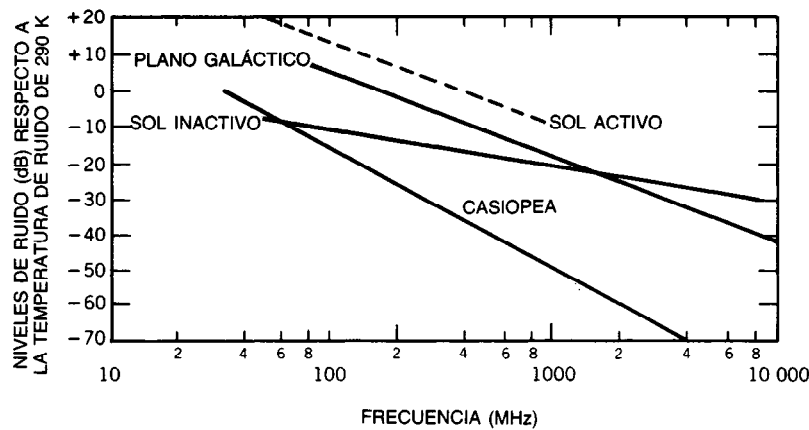


Fig. 47-7. Niveles de ruido galáctico para una antena receptora dipolar de media onda.

experimentado en una región urbana típica. En sitios aislados tranquilos, el nivel de ruido de las fuentes humanas suele ser más bajo que el ruido galáctico en el intervalo de frecuencia de más de 10 MHz. La propagación del ruido de origen humano ocurre principalmente por transmisión en líneas eléctricas y por ondas terrestres; sin embargo, también se verifica por reflexión ionosférica a frecuencias menores de aproximadamente 20 MHz. Las mediciones realizadas indican que el nivel pico de ruido de origen humano no siempre es proporcional al ancho de banda cuando este último es mayor de unos 10 kHz. Según la información más confiable de que se dispone, las máximas intensidades de campo del ruido producido por el hombre (excepto la diatermia y otros ruidos de banda estrecha) aumentan con el ancho de banda del receptor, en especial para anchos de banda mayores de 10 kHz (fig. 47-8).

Fuentes internas

Ruido térmico. El ruido térmico está causado por el movimiento aleatorio de los electrones libres en un conductor excitado por agitación térmica. Todos los componentes pasivos, como guías de onda, cables coaxiales, acopladores, transductores ortomodales, etc., del

sistema receptor generan ruido térmico debido a una pérdida disipativa (pérdida óhmica). Sea R la componente resistiva en ohms de una impedancia Z . El valor cuadrático medio del voltaje por ruido térmico será:

$$E^2 = 4RkTB_n$$

donde k = constante de Boltzmann (1.38×10^{-23} J/K)

T = temperatura absoluta, K

B_n = ancho de banda, Hz

E = voltaje eficaz (rms) por ruido

En esta ecuación se supone que el ruido térmico tiene una distribución uniforme de potencia en todo el ancho de banda B_n . En caso de que dos impedancias Z_1 y Z_2 con componentes resistivos R_1 y R_2 estén en serie a la misma temperatura, el cuadrado del voltaje eficaz resultante es la suma de los cuadrados de los voltajes eficaces por ruido generados en Z_1 y Z_2 :

$$E^2 = E_1^2 + E_2^2 = 4(R_1 + R_2)kTb_n$$

En caso de que las mismas impedancias estén en paralelo a la misma temperatura, la impedancia resultante Z se calcula de la forma acostumbrada para circuitos de corriente alterna, y entonces se determina la

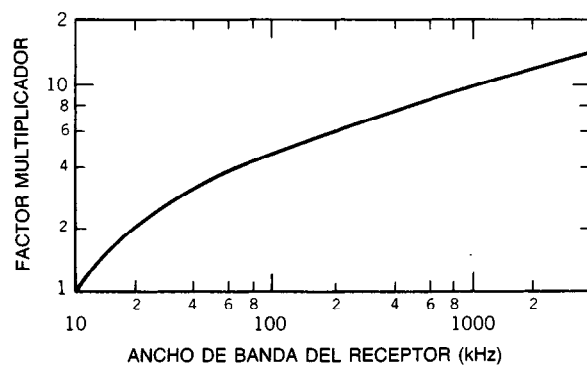


Fig. 47-8. Factor de ancho de banda para el ruido de origen humano.



Fig. 47-9. Transductor ruidoso equivalente.

componente resistiva R de Z . El voltaje eficaz por ruido es el mismo que en una resistencia pura R . En climas templados se acostumbra asignar a T un valor como $1.38 T = 400$, que corresponde aproximadamente a 17°C o 63°F . Así, $E^2 = 1.6 \times 10^{-20} RB_n$.

Ruido de perdigoneo. Está causado por el arribo aleatorio y discreto de electrones a un amplificador. También puede deberse al flujo de corriente en un circuito que incluye un diodo termoiónico o una unión de semiconductor. En tal circuito cada portador de carga contribuye al flujo de corriente durante el tiempo en que viaja del cátodo al ánodo o cruza la unión. La corriente de unión promedio determina el intervalo promedio que transcurre entre la entrada de dos portadores sucesivos a la unión. Sin embargo, el intervalo exacto está sometido a fluctuaciones. Esta aleatoriedad produce un tipo de ruido llamado de perdigoneo, que consiste en pulsos de forma idéntica, pero con amplitudes y tiempos de ocurrencia aleatorios. El valor cuadrático medio de la corriente por ruido de perdigoneo es:

$$I^2 = 2eI_{CD}B_n$$

donde e es la carga de un electrón ($= 1.6008 \times 10^{-19}$ C) e I_{CD} es la corriente (directa) promedio.

47.2.2 Parámetros y características

Factor de ruido

El factor de ruido suele emplearse para describir la calidad de un receptor. En un transductor lineal de dos puertas, el factor de ruido F se define como:

$$F = \frac{S_i/N_i}{S_o/N_o} = \frac{N_o}{kTB_nG}$$

donde S_i = potencia disponible de la señal de entrada
 N_i = potencia disponible del ruido de entrada $= kTB_n$

S_o = potencia disponible de la señal de salida

N_o = potencia disponible del ruido de salida

B_n = ancho de banda equivalente del ruido del receptor

G = ganancia promedio de potencia

Por definición, la potencia del ruido de entrada es equivalente a la potencia del ruido térmico aportada por un resistor conectado a las terminales de entrada del transductor, a una temperatura de $T_0 = 290$ K, y, por tanto, es igual a kT_0B_n . Esta cantidad se denomina a menudo potencia teórica del ruido ambiental. El factor de ruido suele expresarse en decibeles, donde $F_{dB} = 10 \log F$.

Temperatura de ruido

En el caso de los receptores ruidosos ($F > 1$), el concepto de factor de ruido es adecuado para describir el rendimiento. Sin embargo, con el advenimiento de receptores de bajo ruido, tales como los masers y amplificadores paramétricos, una medida más útil del ruido del receptor es la temperatura efectiva del ruido de entrada del receptor. Ésta se define (para un transductor de dos puertas) como la temperatura T_r de un resistor pasivo ficticio, colocado a la entrada de un receptor ideal libre de ruido, que generaría la misma potencia de ruido de salida que el transductor real conectado a una terminal libre de ruido, como se ilustra en la figura 47-9, donde $\Delta N = kT_rB_nG$ y $T_r = (F - 1)T_0 = (F - 1)290$ K. En la figura 47-10 se grafica la relación entre temperatura de ruido y factor de ruido.

Ruido en redes en cascada

Por sencillez, considérense dos circuitos (redes) en cascada, cada uno con el mismo ancho de banda de ruido B_n pero con diferentes factores de ruido (o temperaturas de ruido) y ganancias disponibles. El factor de ruido global F_{12} de las dos redes en cascada es

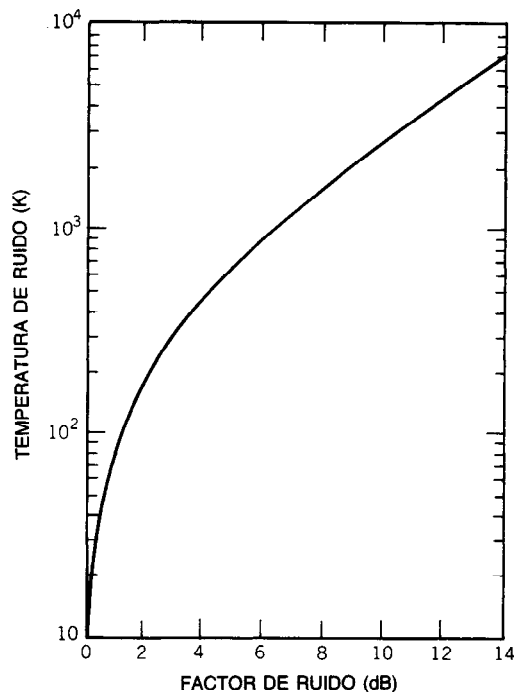


Fig. 47-10. Relación entre temperatura de ruido y factor de ruido.

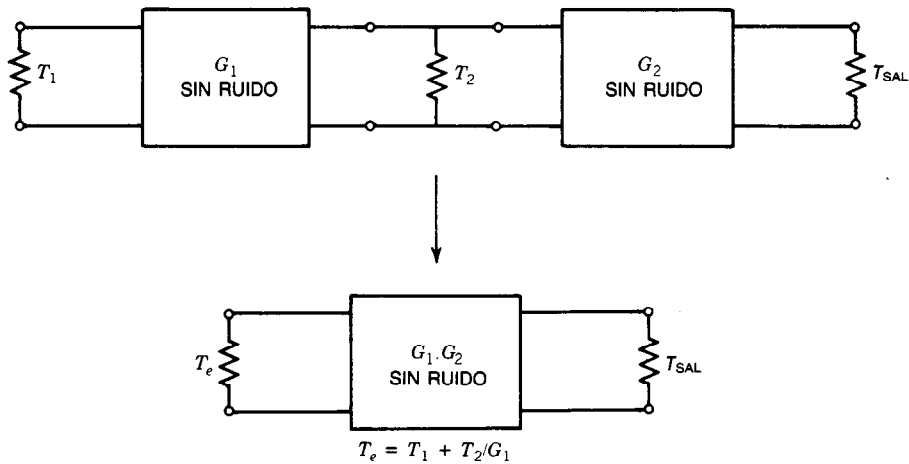


Fig. 47-11. Dos amplificadores en cascada (temperatura de ruido equivalente).

$F_{12} = F_1 + (F_2 - 1)/G_1$, y la temperatura equivalente de ruido de entrada de la combinación es $T_e = T_1 + T_2/G_1$, donde F_1 , G_1 y T_1 son el factor de ruido, ganancia disponible y temperatura de ruido, respectivamente, de la primera red, y F_2 , G_2 y T_2 son los mismos parámetros para la segunda red, como se muestra en la figura 47-11.

De modo similar, el factor de ruido y la temperatura equivalente de ruido de entrada de n redes en cascada están dados por:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_n - 1}{G_1 G_2 \dots G_{n-1}}$$

y

$$T_e = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \dots + \frac{T_n}{G_1 G_2 \dots G_{n-1}}$$

donde $T_i = (F_i - 1)290$ K.

Se observa así que, si la ganancia G_1 de la primera etapa es alta, el factor de ruido o temperatura de ruido globales son prácticamente iguales al factor de ruido F_1 de la primera etapa. En la figura 47-12 se presenta F como función de F_1 , F_2 y G_1 para una red de dos etapas.

Ruido debido a redes con pérdidas

Considérese una red de atenuación con una temperatura física (ambiente) T_p y factor de pérdida $L (> 1)$, como se muestra en la figura 47-13. Supóngase que la red está acoplada en ambos lados con su impedancia

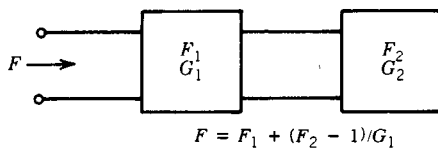


Fig. 47-12. Dos amplificadores en cascada.

característica. La temperatura efectiva de ruido de la entrada del atenuador está dada por $T_e = (L - 1) T_p$ y el factor de ruido es $F = 1 + (L - 1) T_p/T_0$, donde $T_0 = 290$ K. F es igual a L siempre que $T_p = T_0$. Puesto que para una red con pérdidas (disipativa) F depende de la temperatura física del componente, esta temperatura no es un concepto especialmente útil. Por este motivo, suele emplearse más bien la temperatura equivalente de ruido de entrada.

Mediciones del factor de ruido y la temperatura de ruido

La medición del factor de ruido de una red de dos puertas suele hacerse de manera indirecta midiendo la potencia total de ruido suministrada por la red. En el método que se ilustra en la figura 47-14 se emplea en la entrada un generador de señales estándar calibrado para medir experimentalmente la función $G(f)$. Entonces se calcula F conforme a $F = N_0/kT_0 \int G(f) df$, donde N_0 es la potencia del ruido de salida en watts (medida con el generador de señales conectado a la entrada y representando la impedancia de la fuente, pero sin generar señal alguna) y $G(f)$ es la ganancia de potencia disponible (medida con el generador de señales y el medidor de potencia a todas las frecuencias esenciales, y a un nivel de señal suficientemente elevado por encima del nivel de ruido, pero por debajo del nivel de saturación de la red). T_0 es la temperatura de ruido estándar de 290 K y k es la constante de Boltzmann. El factor de ruido también puede calcularse hallando la temperatura de ruido efectiva T_E medida. La medición de T_E para una red de dos puertas suele hacerse por el

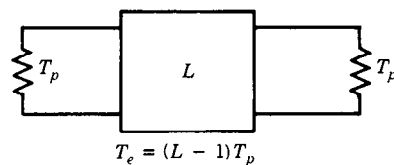


Fig. 47-13. Ruido generado en una red con pérdidas.

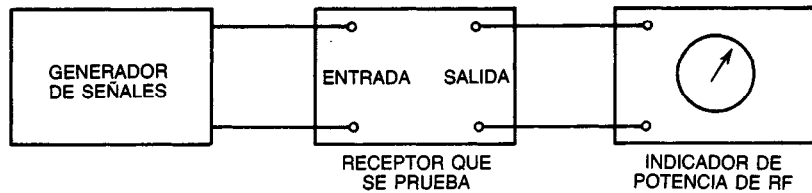


Fig. 47-14. Medición del factor de ruido.

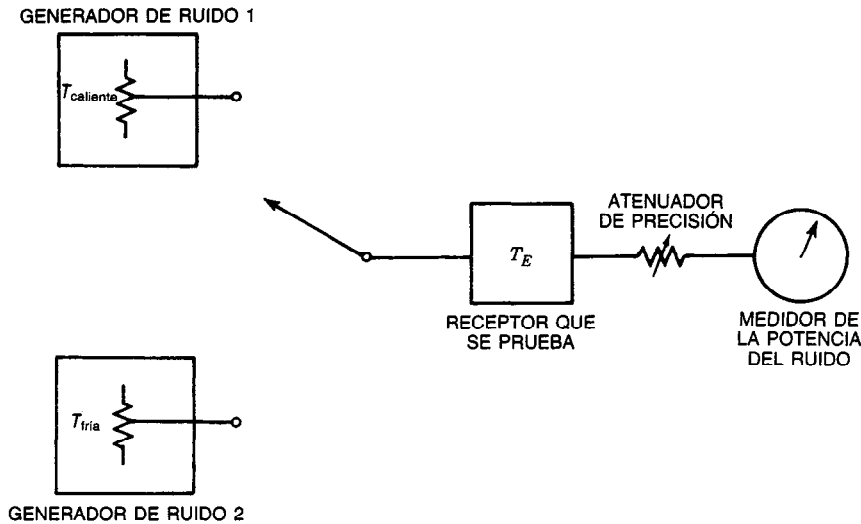


Fig. 47-15. Medición de la temperatura de ruido efectiva en un receptor real.

método del "factor Y". En la figura 47-15 se ilustra el procedimiento para medir la temperatura efectiva de ruido de entrada y el factor de ruido. Como se observa en esa figura, se requieren dos fuentes calibradas de ruido de valor conocido ("caliente" y "fría"). Conmutando entre estas dos fuentes es posible determinar un factor Y mediante un atenuador de precisión y un medidor de potencia del ruido. El factor Y es la razón de la potencia de ruido suministrada por la fuente caliente sobre la suministrada por la fuente fría; esto es:

$$Y = \frac{T_{\text{caliente}} + T_E}{T_{\text{fría}} + T_E}$$

donde T_{caliente} y $T_{\text{fría}}$ son las temperaturas efectivas de ruido de las fuentes caliente y fría, respectivamente, y T_E es la temperatura efectiva de ruido de la red que se prueba. La medición se realiza como sigue:

1. Se conecta el generador caliente de ruido a la red y se ajusta la ganancia para obtener una referencia conveniente en el medidor.
2. Se conmuta el generador frío y se ajusta el atenuador para la misma referencia.
3. Se observa el cambio en la posición de ajuste del atenuador, ΔA (dB).
4. Se calcula $Y = 10^{-\Delta A/10}$.

5. Se calcula el factor de ruido o la temperatura efectiva de ruido de entrada, dados por:

$$T_E = \frac{T_{\text{caliente}} + YT_{\text{fría}}}{Y - 1}$$

y:

$$F = \frac{T_{\text{caliente}}/T_0 - YT_{\text{fría}}/T_0}{Y - 1} + 1$$

Temperatura de ruido del sistema. La potencia total de ruido del sistema receptor puede representarse como una temperatura del sistema T_s de tal nivel que la potencia total de ruido disponible respecto a la terminal de entrada del receptor sea kT_sB_n , donde B_n denota el ancho de banda de ruido del receptor. La temperatura efectiva total de ruido del sistema se compone de:

1. Temperatura efectiva de ruido del espacio ambiental T_a , que es la contribución de la potencia de ruido recibida por la antena de fuentes radiantes externas.
2. Temperatura efectiva de ruido por equipo pasivo, que es el ruido térmico generado por las pérdidas disipativas en el sistema de la línea de transmisión que conecta la antena al receptor.

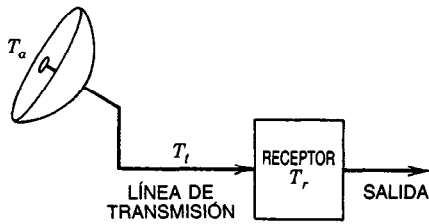


Fig. 47-16. Sistema receptor típico.

3. Temperatura efectiva de ruido por amplificación de RF, o sea, el ruido de fuentes internas al receptor en sí.

Debe hacerse notar en este punto que la división del sistema receptor en tres componentes es arbitraria, puesto que muy a menudo el sistema de antena contiene la línea de transmisión, con lo que se elimina la segunda componente. La temperatura efectiva de ruido de cada una de las tres fuentes de ruido que se esquematizan en la figura 47-16 se representa por T_a , T_i y T_r , respectivamente.

La temperatura de ruido del sistema T_s , referida a la terminal de entrada del receptor está dada por:

$$T_s = \frac{T_a}{L} + T_i \left(1 - \frac{1}{L}\right) + T_r$$

donde L es el factor de pérdida ($L > 1$) de la línea de transmisión y $T_i = 290$ K es la temperatura de referencia para la medición del factor de ruido del receptor. Obsérvese que la temperatura de ruido del sistema referida a la entrada de la línea de transmisión está dada por:

$$T'_s = T_a + (L - 1)T_i + LT_r$$

La importancia de minimizar la pérdida entre la antena y la entrada al receptor se ilustra en la figura 47-17, que contiene una gráfica de la temperatura de ruido del sistema T_s en función de la temperatura de ruido del receptor T_r para una temperatura de la antena de $T_a = 100$ K y diversos valores de L (en decibelios). En la figura 47-18 se presenta la temperatura de ruido del sistema para un sistema receptor típico.

Temperatura de ruido de la antena. La antena recibe ruido que puede originarse en fuentes extraatmosféricas (celestes) o atmosféricas, en el suelo u objetos cercanos al suelo, y en fuentes de origen humano, como la interferencia por ignición de motores. Esta última puede tener características peculiares y no está siempre presente, pero las otras fuentes suelen estarlo, como se muestra en la figura 47-19. La potencia de ruido recibida por una antena de todas las fuentes externas naturales por lo común se representa por medio de una temperatura efectiva de ruido T_a definida por:

$$T_a = \frac{P_a}{kB_n}$$

donde P_a es la potencia de ruido disponible de la antena en el intervalo de frecuencia B_n . La temperatura de ruido de la antena es aquella a la cual debe mantenerse la impedancia de radiación equivalente de la antena para generar la cantidad de potencia de ruido térmico recibida por la antena desde sus alrededores. De este modo, la temperatura de la antena constituye una medida de la cantidad de potencia de ruido recibida con la señal y, por tanto, pasa a ser una limitación fundamental a la sensibilidad en todo sistema de comunicación.

Debe hacerse hincapié en este punto, que la temperatura de ruido de la antena es una temperatura de ruido equivalente o ficticia igual a la de la antena. Así, la temperatura de ruido de la antena T_a es la temperatura que la antena "ve" y no la que "siente".

Por ejemplo, la estructura de la antena puede estar en un ambiente frío de 100 K, y aun así T_a puede ser de varios miles de K. Por otra parte, la estructura de la antena puede medir 300 K, mientras que T_a puede ser de apenas 30 K. La temperatura de ruido de la antena T_a depende tanto de la dirección de la antena como de sus características físicas. La temperatura efectiva de ruido de la antena T_a es entonces la suma de las contribuciones al ruido de distintas fuentes: línea de transmisión que alimenta la antena, pérdidas resistivas en ésta, ruido celeste de la galaxia, el Sol o la Luna, absorción por gases atmosféricos y precipitación, radiación terrestre hacia los lóbulos posteriores de la antena, e interferencia de fuentes de radio humanas. La supresión de los niveles lobulares laterales de la antena en

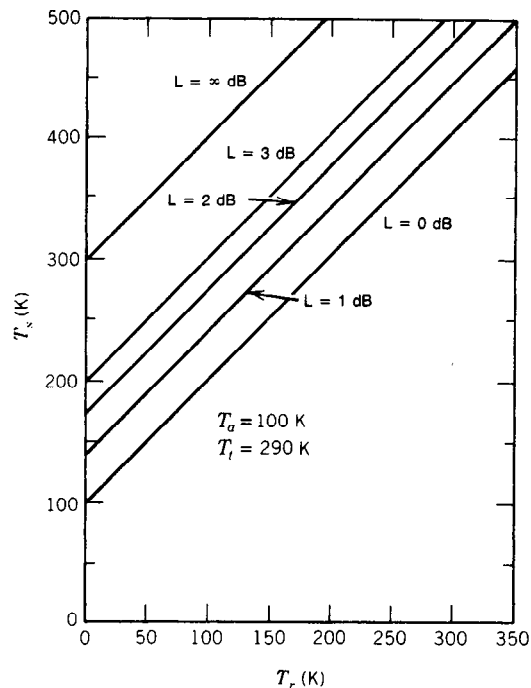


Fig. 47-17. Temperatura de ruido del sistema en comparación con la del receptor.

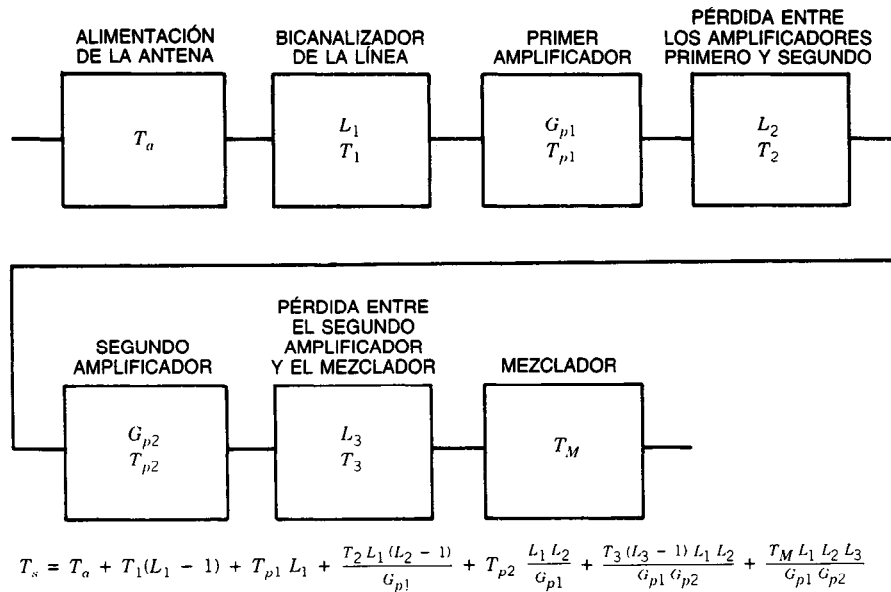


Fig. 47-18. Temperatura de ruido de un sistema receptor típico.

sistemas de bajo ruido es de gran importancia debido a que, mientras el haz principal puede estar dirigido hacia una región fría del cielo, uno de los lóbulos laterales puede estarlo hacia una región caliente, como el Sol, produciendo un valor elevado de T_a . En la figura 47-20 se ilustra el diagrama equivalente de la trayectoria completa para el ruido total de la antenna. Con respecto a la puerta de referencia, todas las fuentes de ruido se combinan en una y la temperatura de ruido correspondiente se calcula como sigue:

$$T_a = \frac{T_{ci}}{L} + \frac{(L - 1)T_{at}}{L} + T_{L1}$$

donde T_{ci} = temperatura de fondo del cielo

L = factor de pérdida del medio ($= e^{\alpha r}$); α es el coeficiente de absorción de la atmósfera

T_{at} = temperatura de la atmósfera

T_{L1} = ruido de la antenna aportado por todos los lóbulos laterales

Al hacer cálculos para el sistema, en muchos casos reviste interés a nivel práctico tomar en cuenta el ruido de la antenna de manera general, sin hacer un cálculo detallado.

En la figura 47-21 se reproducen curvas de ruido de antenas directivas típicas, deducidas por Blake. La exactitud de estas curvas es adecuada en muchas aplicaciones. En su construcción se supone que la antenna está plantada en tierra, y no son aplicables a antenas aéreas de radar ni a antenas plantadas en tierra a altitudes extremas. El resultado de esta suposición es que se incluye el efecto de toda la columna atmosférica comprendida entre la antenna y las fuentes de ruido extraterrestres. Las curvas discontinuas representan el ruido cósmico y atmosférico máximo y mínimo.

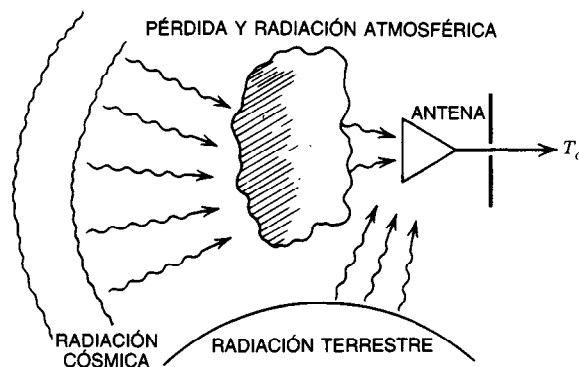


Fig. 47-19. Elementos externos que contribuyen a la temperatura de ruido de una antenna.

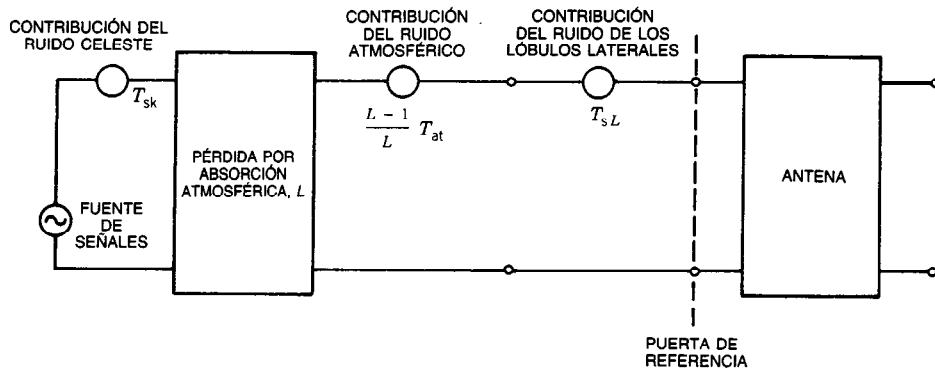


Fig. 47-20. Circuito equivalente para el cálculo de la temperatura de ruido de una antena.

Medición del ruido de la antena

La temperatura efectiva de ruido de la antena T_a también puede medirse por el método del factor Y empleando una fuente de ruido calibrada de valor conocido, como se observa en la figura 47-22:

$$Y = \frac{T_p + T_E}{T_a + T_E}$$

y:

$$T_a = \frac{T_p + T_E}{Y - T_E}$$

donde T_p es la temperatura efectiva de ruido de la superficie de ruido calibrada y T_E es la temperatura efectiva de ruido del receptor medida por el método del factor Y .

Ruido gaussiano blanco aditivo

El ruido térmico tiene una densidad espectral de potencia bastante uniforme hasta frecuencias del orden de los 10^{13} Hz, y para el ruido de perdigoneo esa densidad es razonablemente constante hasta frecuencias del orden del recíproco del tiempo de tránsito de los portadores de carga de un lado a otro de la unión. De modo similar, otras fuentes de ruido, como la atmosférica y la galáctica, tienen intervalos espectrales muy amplios. Estos tipos de ruido suelen denominarse ruido gaussiano blanco aditivo (AWGN, de *additive white gaussian noise*). El **ruido blanco** es aquel cuya densidad espectral de potencia es uniforme en todo el intervalo de frecuencia de interés. Se usa el término "blanco" por analogía con la luz blanca, que es una superposición de todas las componentes espectrales visibles. El término "aditivo" se refiere al ruido que se suma a la señal que recibirá el sistema de comunicación, y "gaussiano" se refiere al ruido cuya función de densidad de probabilidad es normal o de Gauss. En casi todos los sistemas de comunicación y en una amplia variedad de circunstancias, la suposición de una función de densidad gaussiana es justificable. Sin embargo, tal suposición no es universalmente válida. Por ejemplo, si se aplica ruido gaussiano a la entrada de un circuito rectificador, la salida no es gaussiana. De modo similar, bien puede

ocurrir que el ruido encontrado en ciertos canales consista en perturbaciones pulsátiles cortas con distribución de amplitud definitivamente no gaussiana. En la figura 47-23 se presenta un modelo típico de AWGN, donde $s(t)$ es la señal, $n(t)$ es el ruido del canal e $y(t)$ es la señal recibida más el ruido. En la figura 47-24 se ilustra el espectro de la señal recibida, donde N_0 es la densidad espectral del ruido (unilateral, constante en todo el intervalo de frecuencia de interés) y depende de la temperatura del sistema receptor (esto es, $N_0 = kT_s$).

Ancho de banda del ruido equivalente

El ancho de banda del ruido equivalente B_n es:

$$B_n = (1/2\pi) \int_{-\infty}^{\infty} \frac{|H(j\omega)|^2 d\omega}{|H(j\omega_0)|^2}$$

donde $H(j\omega)$ es la función de transferencia de la red y ω_0 denota la frecuencia angular de la respuesta máxima. Por tanto, B_n es igual al ancho de banda de un filtro rectangular equivalente de altura $H(j\omega_0)$, cuya potencia de salida de ruido es igual al ruido total de salida de la red.

Efecto del filtrado

Filtro pasabajas rectangular (ideal)

$$P_n = N_0 B_n$$

donde P_n es la potencia del ruido de salida.

Filtro pasabajas RC

$$P_n = \frac{\pi N_0 f_c}{2}$$

donde f_c es la frecuencia a 3 dB del filtro pasabajas.

Filtro pasabanda rectangular

$$P_n = N_0(f_2 - f_1)$$

donde $(f_2 - f_1)$ es el ancho de banda del filtro.

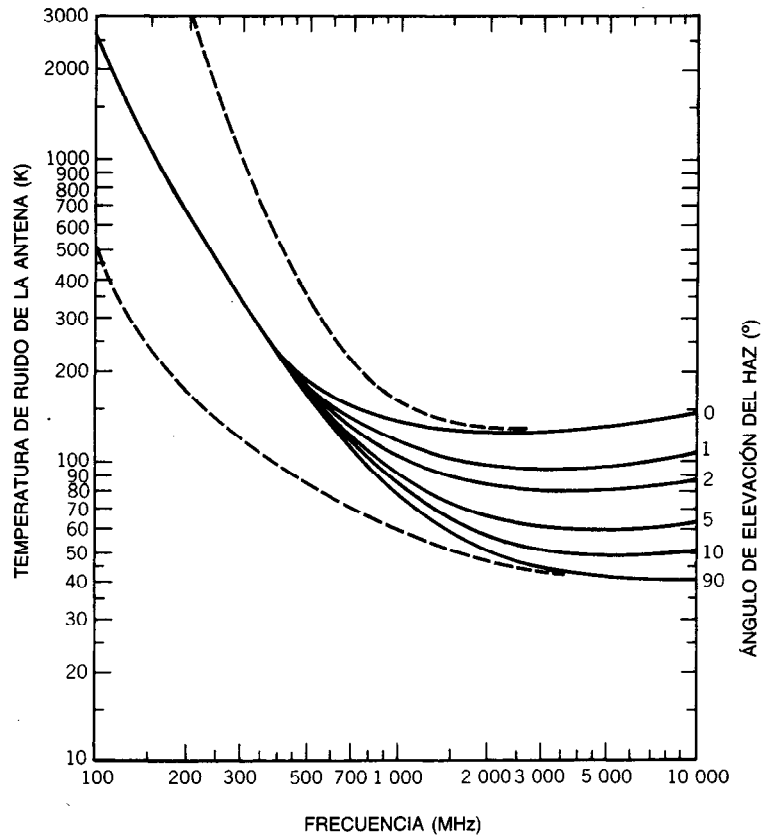


Fig. 47-21. Temperatura de ruido de una atena direccional para condiciones ambientales representativas.

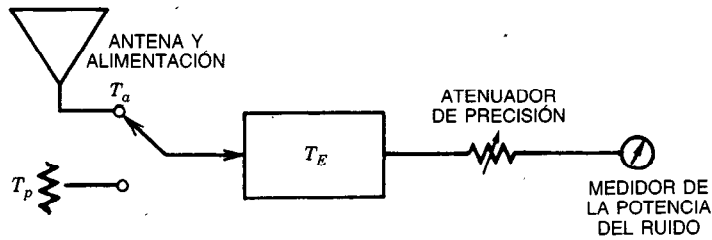


Fig. 47-22. Medición de la temperatura efectiva de ruido de la fuente.

Filtro diferenciador (seguido por un filtro pasabajas rectangular)

$$P_n = \frac{4\pi^2 N_0 \tau^2 B_n^3}{3}$$

donde τ es un factor constante de proporcionalidad.

Integrador

$$P_n = \frac{N_0 T}{2\tau^2}$$

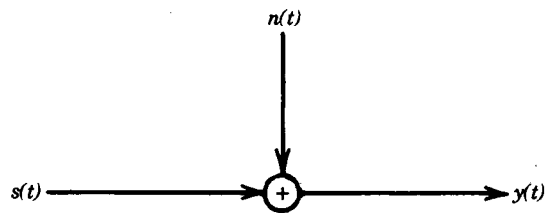


Fig. 47-23. Modelo del ruido gaussiano blanco aditivo (AWGN).

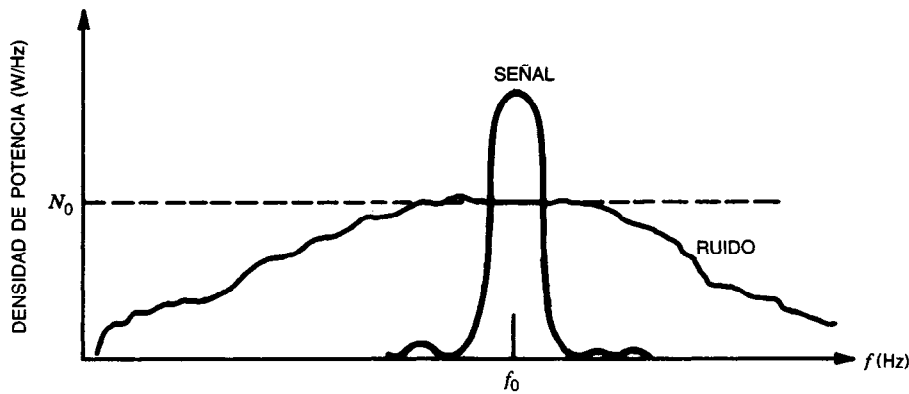


Fig. 47-24. Espectro de la señal recibida.

donde T es el tiempo de integración y τ representa a un factor constante de proporcionalidad.

Representación del ruido de paso de banda

Para minimizar la potencia de ruido que llega al demodulador de un sistema receptor, suele emplearse un filtro antes de aquel elemento. El ancho de banda B_n se hace tan estrecho como sea posible a fin de evitar transmitir cualquier ruido innecesario. La salida de este filtro pasabanda suele representarse por:

$$n(t) = n_c(t)\cos 2\pi f_0 t + n_s(t)\sin 2\pi f_0 t$$

donde f_0 es una frecuencia arbitraria. A menudo conviene emplear esta representación al tratar ruido limitado a una banda de frecuencia relativamente estrecha en la vecindad de f_0 . Por esta causa, esta ecuación suele llamarse representación de banda estrecha. También se usa el término "representación de componentes en cuadratura" debido a la presencia en la ecuación de sinusoides en cuadratura. Los términos $n_c(t)$ y $n_s(t)$ son señales de ruido de paso bajo limitadas en banda a B_n . Como $n(t)$, $n_c(t)$ y $n_s(t)$ son procesos aleatorios gaussianos. Las funciones densidad de probabilidad de las variables aleatorias n , n_c y n_s [esto es, $n(t)$, $n_c(t)$ y $n_s(t)$ en cualquier tiempo fijo] están dadas por:

$$P(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{n^2}{2\sigma_n^2}\right)$$

y:

$$P(n_c) = \frac{1}{\sqrt{2\pi\sigma_{n_c}^2}} \exp\left(-\frac{n_c^2}{2\sigma_{n_c}^2}\right)$$

$$P(n_s) = \frac{1}{\sqrt{2\pi\sigma_{n_s}^2}} \exp\left(-\frac{n_s^2}{2\sigma_{n_s}^2}\right)$$

donde $\sigma_n^2 = \sigma_{n_c}^2 = \sigma_{n_s}^2 = N_0 B_n$ = potencia total de ruido = valores cuadráticos medios de $n(t)$, $n_c(t)$ y $n_s(t)$, respectivamente. También es posible representar $n(t)$ en forma polar:

$$n(t) = R(t)\cos[2\pi f_0 t + \theta(t)]$$

donde:

$$R(t) = \sqrt{n_c^2(t) + n_s^2(t)}$$

y:

$$\theta(t) = \tan^{-1}\left[\frac{n_s(t)}{n_c(t)}\right]$$

Puesto que tanto $n_c(t)$ como $n_s(t)$ son señales que varían muy lentamente, se concluye que $R(t)$ y $\theta(t)$ son también señales que varían muy lentamente. La señal de ruido de paso de banda con espectro de banda estrecha tiene aspecto de señal sinusoidal cuya amplitud y fase varían lentamente. La envolvente de esta señal está dada por $R(t)$, y la fase es $\theta(t)$.

LIBROS Y REVISTAS

- A. Benoit, "Signal Attenuation Due to Neutral Oxygen and Water Vapour, Rain and Clouds", *Microwave J.*, pp. 73-80.
- L. V. Blake, "Antenna and Receiving-System Noise-Temperature Calculation", *NRL Rep. 5668*.
- CCIR 10.^a Plenary Assembly, "World Distribution and Characteristics of Atmospheric Radio Waves", Informe n.º 322, Ginebra.
- CCIR 11.^a Plenary Assembly, vol. 2, Informe n.º 234, Oslo.
- R. K. Crane, "Propagation Phenomena Affecting Satellite Communication Systems", *Proc. IEEE*, p. 178.
- W. Q. Crichlow y colaboradores, "Special Report on Characteristics of Terrestrial Radio Noise", International Scientific Radio Union (URSI), Comisión IV.
- J. Dijk, M. Jeuken y F. J. Maanders, "Antenna Noise Temperature", *Proc. IEE*, **115**(10): 1403-1410.
- Fredrick Research Corp., *Handbook of Radio-Frequency Interference*, vols. 1 a 4, Wheaton, MA.

- H. H. Grimm, "Noise Temperature in Passive Circuits", *Microwave J.*, pp. 52-54.
- D. C. Hogg, "Millimeter-Wave Communication Through the Atmosphere", *Science* **159** (3810):39-46.
- J. B. Johnson, "Thermal Agitation of Electricity in Conductors", *Phys. Rev.* **32**:97-109.
- D. K. C. MacDonald, *Noise and Fluctuations: An Introduction*, Wiley, Nueva York.
- W. W. Mumford y E. H. Scheibe, *Noise Performance Factors in Communication Systems*, Horizon House-Microwave, Delham, MA.
- J. L. Powsey y R. N. Bracewell, *Radio Astronomy*, Clarendon Press, Oxford, Gran Bretaña.
- W. V. T. Rusch y P. D. Potter, *Analysis of Reflector Antennas*, Academic, Nueva York.
- H. Staras, "The Propagation of Wideband Signals Through the Ionosphere", *Proc. IRE* **49**:1211.
- J. L. Steinberg y J. Lequeux, *Radio Astronomy*, McGraw-Hill, Nueva York.
- R. D. Stephenson, External Noise, en A. V. Balakrishnan, ed., *Space Communications*, cap. 6, McGraw-Hill, Nueva York.
- H. Taub y D. L. Schilling, *Principles of Communication Systems*, Chs. 7, 8, 9, McGraw-Hill, Nueva York.
- A. D. Watt, R. M. Coon, E. L. Maxwell y R. W. Plush, "Performance of Some Radio Systems in the Presence of Thermal and Atmospheric Noise", *Proc. IRE* **46**:1914-1923.

47.3 TEORÍA DE LA INFORMACIÓN*

Rodger E. Ziemer y William H. Tranter

Aquí se presenta la teoría de la información, haciendo uso de las características de funcionamiento de un sistema de comunicación *ideal* u óptimo. Tal funcionamiento constituye una base significativa para comparar el de sistemas realizables. Esto ilustra la ganancia en rendimiento que puede obtenerse implementando esquemas de transmisión y detección más complicados.

El principal estímulo para el estudio de la teoría de la información proviene del teorema de la codificación de Shannon, algunas veces llamado segundo teorema de Shannon, el cual puede plantearse como sigue: Si una fuente tiene una velocidad de transmisión de información menor que la capacidad del canal, existe un procedimiento de codificación tal que la salida de la fuente puede transmitirse por ese canal con una probabilidad de error arbitrariamente pequeña.

Este enunciado es sorprendente. Parece que se desea dar a entender que la transmisión y recepción pueden realizarse con un error *despreciable*, incluso en presencia de ruido. Para comprender el proceso de la codificación y su efecto en el diseño y funcionamiento de los sistemas de comunicación es necesario entender algunos conceptos de la teoría de la información.

* Reproducido con permiso de R. E. Ziemer y W. H. Tranter, *Principles of Communications*, Houghton Mifflin, Boston.

47.3.1 Conceptos básicos

Considérese una situación hipotética en un salón de clases que se presenta al final de un periodo escolar. El profesor hace una de las siguientes declaraciones al grupo:

- Los veré el próximo periodo de clases. (Declaración A.)
- Mi colega dará clases el próximo periodo. (Declaración B.)
- Todos obtuvieron "10" en el curso, y no habrá más exámenes. (Declaración C.)

¿Cuál es la información relativa proporcionada a los estudiantes por cada una de estas declaraciones? Obviamente la declaración A aporta poca información, puesto que el grupo normalmente supondrá que su profesor habitual seguirá con ellos el próximo curso; esto es, la probabilidad de que el profesor habitual dé clases, $P(A)$, es casi la unidad. Intuitivamente, se sabe que el enunciado B contiene más información, y la probabilidad de que otro profesor dé clases el siguiente periodo, $P(B)$, es relativamente baja. El tercer enunciado, C, contiene una gran cantidad de información para todo el grupo, y todo el mundo estará de acuerdo en que tal declaración tiene muy poca probabilidad de presentarse en una situación escolar típica. Se observa que cuanto menor sea la probabilidad de un enunciado, mayor es la información aportada por él. Dicho de otra manera, la sorpresa de los estudiantes al escuchar una declaración parece ser una buena medida de la cantidad de información contenida en tal declaración.

A continuación se define matemáticamente la información de una forma consistente.

Información

Sea x_j un evento (suceso) que ocurre con probabilidad $p(x_j)$. Si se afirma que ha ocurrido el evento x_j , entonces se dirá que se han recibido:

$$I(x_j) = \log_a \frac{1}{p(x_j)} = -\log_a p(x_j) \quad (47-1)$$

unidades de información. Esta definición de la información es consistente con el ejemplo anterior, puesto que $I(x_j)$ aumenta de manera inversamente proporcional con respecto a $p(x_j)$.

La base del logaritmo incluido en la ecuación 47-1 es arbitraria y determina las unidades en las cuales se mide la información. Hartley, quien sugirió por primera vez en 1928 el uso de la medida logarítmica de información, empleó los logaritmos de base 10, y la medida de información fue denominada hartley. En la actualidad la norma es el uso de los logaritmos de base 2, y la unidad de información es la unidad binaria o bit. Algunas veces se emplean logaritmos de base e , en cuyo caso la unidad correspondiente es el nat.

Existen muchas razones para usar consistentemente el logaritmo de base 2 al medir información. El experimento aleatorio más simple que uno puede imaginarse es aquel con dos resultados igualmente probables,

como el de arrojar al aire una moneda no defectuosa. El conocimiento de cada resultado tiene asociado con él un bit de información. Además, dado que la computadora digital es una máquina binaria, cada 0 y cada 1 lógicos tienen asociados un bit de información, suponiendo que estos estados lógicos son igualmente probables.

Ejemplo 47.1. Considérese un experimento aleatorio con 16 resultados igualmente probables. La información asociada con cada resultado es:

$$I(x_j) = -\log_2 \frac{1}{16} = \log_2 16 = 4 \text{ bits}$$

donde j varía de 1 a 16. La información es mayor que 1 bit, puesto que la probabilidad de cada resultado es mucho menor que un medio.

Entropía

En general, la *información promedio* asociada con el resultado de un experimento reviste mayor interés que la asociada con cada evento particular. La información promedio asociada con una variable aleatoria discreta X se define como la entropía $H(X)$. De esta forma:

$$H(X) = E\{I(x_j)\} = - \sum_{j=1}^n p(x_j) \log_2 p(x_j) \quad (47-2)$$

donde n es el número total de posibles resultados. La entropía puede considerarse una incertidumbre promedio, y por tanto debe ser máxima cuando todos los resultados son igualmente probables.

Ejemplo 47.2. Para una fuente binaria $p(1) = \alpha$ y $p(0) = 1 - \alpha = \beta$, deducir la entropía de la fuente como función de α y graficar $H(\alpha)$ en función de α en el intervalo $0 \leq \alpha \leq 1$.

De la ecuación 47-2:

$$H(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha) \quad (47-3)$$

Esto se grafica en la figura 47-25. Debe hacerse notar el máximo. Si $\alpha = 1/2$, ambos símbolos tienen la misma probabilidad y la incertidumbre es máxima. Si $\alpha \neq 1/2$, un símbolo tiene mayor probabilidad de ocurrencia que el otro, y hay menor incertidumbre en cuanto a cuál de ellos se presentará en la salida de la fuente. Si α o β es igual a cero, la incertidumbre es cero, puesto que se sabe qué símbolo se presentará.

En base al ejemplo anterior se concluye, intuitivamente, que la función entropía tiene un máximo, y que ese máximo ocurre cuando ambas probabilidades son iguales. Este hecho reviste suficiente importancia para autorizar una deducción más completa.

Supóngase que un proceso aleatorio tiene n resultados posibles y que p_n es una variable que depende de las otras probabilidades. Así:

$$p_n = 1 - (p_1 + p_2 + \dots + p_k + \dots + p_{n-1}) \quad (47-4)$$

donde p_k es $p(x_j)$ en notación concisa. La entropía asociada con el proceso es:

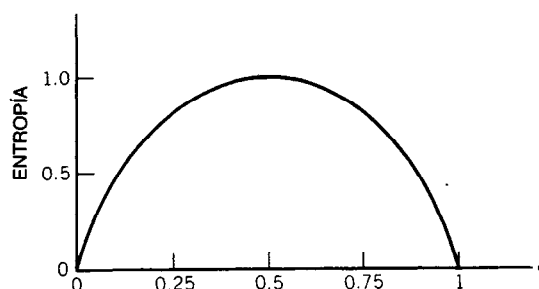


Fig. 47-25. Entropía de una fuente binaria.

$$H = - \sum_{i=1}^n p_i \log_2 p_i \quad (47-5)$$

Para encontrar el valor máximo de la entropía ésta se diferencia con respecto a p_k , manteniendo constantes todas las probabilidades excepto p_k y p_n . Esto produce una relación entre estas últimas variables, la cual indica el valor máximo de H . Puesto que todas las derivadas son cero excepto aquellas en las que se encuentran p_k y p_n :

$$\frac{dH}{dp_k} = \frac{d}{dp_k} (-p_k \log_2 p_k - p_n \log_2 p_n) \quad (47-6)$$

Empleando la ecuación 47-4 y:

$$\frac{d}{dx} \log_a u = \frac{1}{u} \log_a e \frac{du}{dx}$$

se obtiene:

$$\begin{aligned} \frac{dH}{dp_k} = & -p_k \frac{1}{p_k} \log_2 e - \log_2 p_k + \\ & + p_n \frac{1}{p_n} \log_2 e + \log_2 p_n \end{aligned}$$

o bien:

$$\frac{dH}{dp_k} = \log_2 \frac{p_n}{p_k}$$

que es cero si $p_k = p_n$. Puesto que p_k es arbitraria:

$$p_1 = p_2 = \dots = p_n = \frac{1}{n} \quad (47-7)$$

Para demostrar que la condición anterior produce un máximo y no un mínimo, obsérvese que cuando $p_1 = 1$ y todas las demás probabilidades son cero la entropía es cero. En base a la ecuación 47-5, en el caso en que todas las probabilidades sean iguales, $H = \log_2 n$.

Representaciones de canales

En la mayor parte de este capítulo se considera que el canal de comunicación está desprovisto de memoria.

En tales canales, la salida en un instante dado es función de la entrada del canal *en ese instante* y no de las entradas anteriores al canal. Los canales discretos sin memoria están completamente especificados por un conjunto de probabilidades condicionales, en el que la probabilidad de cada estado de salida se relaciona con las probabilidades de la entrada. En la figura 47-26 se presenta el diagrama de un canal con dos entradas y tres salidas. Cada trayectoria posible de entrada a salida se indica junto con una probabilidad condicional p_{ij} , equivalente en notación concisa a $p(y_j | x_i)$. De este modo, p_{ij} es la probabilidad condicional de obtener la salida y_j , dado que la entrada es x_i , y se denomina probabilidad de transición del canal.

Puede observarse en la figura 47-26 que el canal está completamente especificado por el conjunto completo de probabilidad de transición. En consecuencia, el canal a menudo se especifica por la matriz de probabilidades de transición $[P(Y | X)]$, donde, para el canal de la figura 47-26:

$$[P(Y | X)] = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & p(y_3|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & p(y_3|x_2) \end{bmatrix} \quad (47-8)$$

Dado que cada entrada al canal produce una salida, cada renglón de la matriz debe sumar la unidad.

La matriz del canal es útil para deducir las probabilidades de salida, dadas las probabilidades de entrada. Por ejemplo, si las probabilidades de entrada $P(X)$ se representan por medio de la matriz renglón:

$$[P(X)] = [p(x_1) \quad p(x_2)] \quad (47-9)$$

entonces:

$$[P(Y)] = [p(y_1) \quad p(y_2) \quad p(y_3)] \quad (47-10)$$

que se calcula por medio de:

$$[P(Y)] = [P(X)][P(Y | X)] \quad (47-11)$$

Si $[P(X)]$ se expresa como una matriz diagonal, la ecuación 47-11 produce una matriz $[P(X, Y)]$. Todos los elementos de la matriz tienen la forma $p(x_i)p(y_j | x_i)$, o $p(x_i, y_j)$. Esta matriz se conoce como matriz de probabilidad conjunta, y el término $p(x_i, y_j)$ es la probabilidad conjunta de transmitir x_i así como recibir y_j .

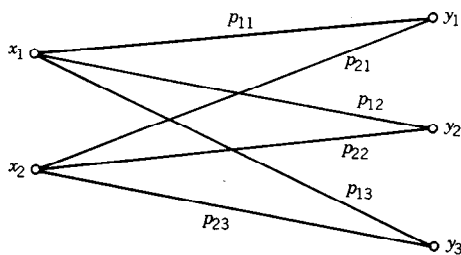


Fig. 47-26. Diagrama de un canal.

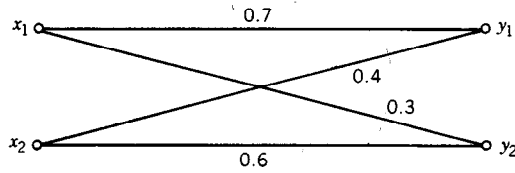


Fig. 47-27. Canal binario.

Ejemplo 47.3. Considérese el canal de entrada-salida binario de la figura 47-27. La matriz de probabilidades de transición es:

$$P[Y|X] = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Si las probabilidades de entrada son $P(x_1) = 0.5$ y $P(x_2) = 0.5$, las probabilidades de salida son:

$$P(Y) = [0.5 \quad 0.5] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [0.55 \quad 0.45]$$

y la matriz de probabilidad conjunta es:

$$P[X, Y] = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.35 & 0.15 \\ 0.2 & 0.3 \end{bmatrix}$$

Entropía conjunta y condicional

Empleando las probabilidades de entrada $p(x_i)$, las probabilidades de salida $p(y_j)$, las probabilidades de transición $p(y_j | x_i)$ y las probabilidades conjuntas $p(x_i, y_j)$ es posible definir algunas funciones entropía distintas para un canal con n entradas y m salidas. Dichas funciones son:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (47-12)$$

$$H(Y) = - \sum_{j=1}^m p(y_j) \log_2 p(y_j) \quad (47-13)$$

$$H(Y|X) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(y_j | x_i) \quad (47-14)$$

y:

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j) \quad (47-15)$$

Otra entropía útil, $H(X | Y)$, algunas veces llamada equivocación, se define como:

$$H(X|Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i | y_j) \quad (47-16)$$

Estas entropías se interpretan con facilidad. $H(X)$ es la incertidumbre promedio de la fuente, mientras que

$H(Y)$ es la incertidumbre promedio del símbolo recibido. De modo similar, $H(X | Y)$ es una medida de la incertidumbre promedio en cuanto al símbolo transmitido después de haber recibido un símbolo. La función $H(Y | X)$ es la incertidumbre promedio del símbolo recibido dado, que transmitió X . La entropía conjunta $H(X, Y)$ es la incertidumbre promedio del sistema de comunicación como un todo.

Dos relaciones importantes y útiles que se pueden obtener directamente de las definiciones de las distintas entropías son:

$$H(X, Y) = H(X | Y) + H(Y) \quad (47-17)$$

y:

$$H(X, Y) = H(Y | X) + H(X) \quad (47-18)$$

Capacidad de un canal

Considérese por un momento un observador a la salida de un canal. La incertidumbre promedio del observador respecto a la entrada del canal tendrá algún valor antes de la recepción de una salida, y por lo general su incertidumbre promedio sobre la entrada disminuirá al recibir la salida. En otras palabras, $H(X | Y) \leq H(X)$. El decremento en la incertidumbre promedio del observador en cuanto a la señal transmitida cuando se recibe la salida es una medida de la información promedio transmitida. Esto se define como transinformación o información mutua $I(X; Y)$. Se tiene entonces que:

$$I(X; Y) = H(X) - H(X | Y) \quad (47-19)$$

Se sigue de las ecuaciones 47-17 y 47-18 que la ecuación 47-19 también puede expresarse de la forma:

$$I(X; Y) = H(Y) - H(Y | X) \quad (47-20)$$

Debe observarse que la transinformación es función de las probabilidades de la fuente, así como de las probabilidades de transición del canal.

La capacidad del canal C se define como el valor máximo de la transinformación, que es la máxima información promedio *por símbolo* que puede transmitirse por el canal. Así:

$$C = \max[I(X; Y)] \quad (47-21)$$

La maximización es con respecto a las probabilidades de la fuente, puesto que las probabilidades de transición son fijadas por el canal. Sin embargo, la capacidad del canal es función sólo de las probabilidades de transición de ese canal, puesto que el proceso de maximización elimina la dependencia respecto a las probabilidades de la fuente.

Ejemplo 47.4. Encontrar la capacidad del canal discreto sin ruido de la figura 47-28.

Se comienza con:

$$I(X; Y) = H(X) - H(X | Y)$$

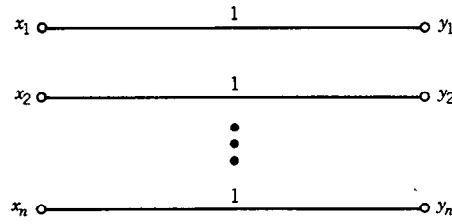


Fig. 47-28. Canal libre de ruido.

y se expresa:

$$H(X | Y) = - \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 p(x_i | y_j)$$

Para el canal sin ruido, todas las $p(x_i, y_j)$ y las $p(x_i | y_j)$ son cero, a menos que $i = j$. Para $i = j$, $p(x_i | y_j)$ es la unidad. Por tanto $H(X | Y)$ es cero para el canal sin ruido, y:

$$I(X; Y) = H(X)$$

Se ha visto que la entropía de una fuente es máxima si todos los símbolos de la fuente son igualmente probables. De esta forma:

$$C = \sum_{i=1}^n \frac{1}{n} \log_2 n = \log_2 n \quad (47-22)$$

Ejemplo 47.5. Encontrar la capacidad del canal simétrico binario de la figura 47-29.

Este problema, por lo general, reviste considerable importancia práctica en el área de las comunicaciones digitales binarias. La capacidad se determinará maximizando:

$$I(X; Y) = H(Y) - H(Y | X)$$

donde:

$$H(Y | X) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log_2 p(y_j | x_i)$$

Empleando las probabilidades definidas en la figura 47-29 se obtiene:

$$\begin{aligned} H(Y | X) = & - \alpha \log_2 p - (1 - \alpha) p \log_2 p \\ & - \alpha q \log_2 q - (1 - \alpha) q \log_2 q \end{aligned}$$

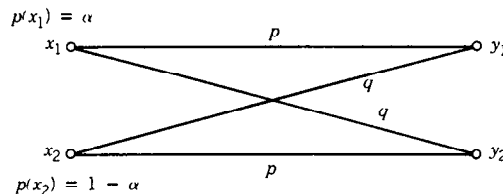


Fig. 47-29. Canal simétrico binario.

o bien:

$$H(Y|X) = -p \log_2 p - q \log_2 q$$

Así:

$$I(X;Y) = H(Y) + p \log_2 p + q \log_2 q$$

que es máxima cuando $H(Y)$ es máxima. Dado que la salida del sistema es binaria, $H(Y)$ es máxima cuando cada entrada tiene probabilidad de $1/2$, y se alcanza cuando las entradas son igualmente probables. En este caso $H(Y)$ es la unidad, y la capacidad del canal es:

$$C = 1 + p \log_2 p + q \log_2 q = 1 - H(p) \quad (47-23)$$

donde $H(p)$ se define como en la ecuación 47-3.

En la figura 47-30 se presenta esquemáticamente la capacidad de un canal simétrico binario. Como sería de esperar, si $p = 0$ o 1 , la salida del canal está completamente determinada por la entrada, y la capacidad es 1 bit por símbolo. Si $p = 1/2$, un símbolo de entrada genera cualquiera de las salidas con igual probabilidad, y la capacidad es cero.

La probabilidad de error P_E de un canal simétrico binario se calcula con facilidad. En base a:

$$P_E = \sum_{i=1}^2 p(e|x_i)p(x_i) \quad (47-24)$$

donde $p(e | x_i)$ es la probabilidad de error dada la entrada x_i , se tiene que:

$$P_E = qp(x_1) + qp(x_2)$$

Así:

$$P_E = q \quad (47-24a)$$

Esta expresión establece que la probabilidad incondicional de error P_E es igual a la probabilidad condicional de error $p(y_j | x_i)$, $i \neq j$.

P_E es una función decreciente de la energía de los símbolos recibidos. Dado que la energía de los símbolos es la potencia recibida multiplicada por el periodo de los símbolos, se sigue que, si la potencia transmitida es fija, es posible reducir la probabilidad de error reduciendo la velocidad de la fuente. Esto puede lograrse eliminando la redundancia en la fuente a través de un proceso denominado codificación de la fuente.

Canal continuo

La capacidad en bits por segundo de un canal continuo con ruido gaussiano blanco aditivo está dada por:

$$C_c = B \log_2 \left(1 + \frac{S}{N} \right) \quad (47-25)$$

donde B es el ancho de banda en hertz y S/N es la razón de señal sobre ruido. Esta expresión particular de la capacidad de un canal se conoce como ley de Shannon-Hartley. El subíndice se usa para distinguir la

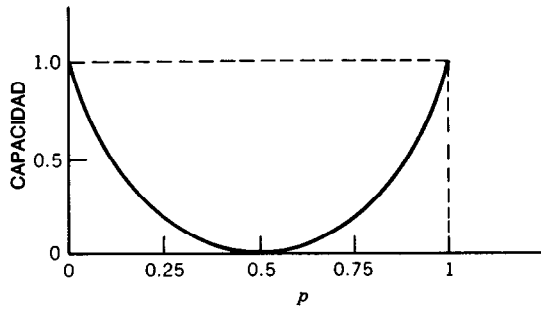


Fig. 47-30. Capacidad de un canal simétrico binario.

ecuación 47-25, que tiene unidades de bits por segundo, de la ecuación 47-21, con unidades de bits por símbolo.

La solución intermedia entre ancho de banda y razón de señal sobre ruido puede determinarse a partir de la ley de Shannon-Hartley. Para el caso sin ruido, con una razón de señal sobre ruido infinita, la capacidad es infinita dado cualquier ancho de banda distinto de cero. Sin embargo, como se demostrará, la capacidad no puede hacerse infinitamente grande incrementando el ancho de banda si existe ruido.

En el caso de un canal con ruido, resulta interesante calcular la razón de señal sobre ruido que aún permite la transmisión a una velocidad igual a la capacidad del canal. En primer lugar la ecuación 47-25 se expresa de la forma:

$$C_c = B \log_2 \left(1 + \frac{S}{N_0 B} \right) \quad (47-26)$$

donde S es la energía de la señal por unidad de tiempo (de modo que representa una potencia) y N_0 es la densidad espectral de potencia del ruido. Es posible determinar una cota útil encontrando el límite cuando B tiende a infinito.

Antes de encontrar el límite es conveniente volver a expresar la ecuación 47-26:

$$C_c = \frac{S}{N_0} \log_2 \left[\left(1 + \frac{S}{N_0 B} \right)^{N_0 B/S} \right] \quad (47-27)$$

Es posible emplear entonces:

$$\lim_{x \rightarrow 0} (1+x)^{1/x} = e$$

para expresar:

$$\lim_{B \rightarrow \infty} C_c = \frac{S}{N_0} \log_2 e \quad (47-28)$$

La potencia de la señal ($S = E/T$) para señalización M -aria puede expresarse en términos de la velocidad si se reconoce que:

$$R = \frac{\log_2 M}{T} \quad (47-29)$$

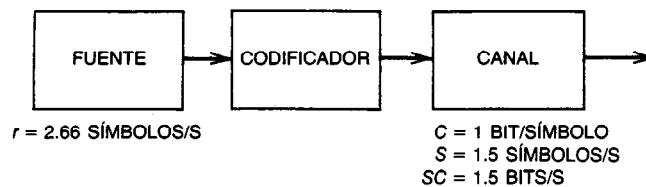


Fig. 47-31. Esquema de transmisión.

si todas las M señales son equiprobables. Por tanto, la potencia de la señal se expresa en la siguiente forma:

$$S = \frac{ER}{\log_2 M} \quad (47-30)$$

y la ecuación 47-28 pasa a ser:

$$\lim_{B \rightarrow \infty} C_c = \frac{ER}{N_0 \log_2 M} \log_2 e \quad (47-31)$$

que para $R = C$ produce:

$$\frac{E}{N_0 \log_2 M} = \frac{1}{\log_2 e} \cong \frac{1}{1.44} \cong -1.6 \text{ dB} \quad (47-32)$$

De esta forma, para $E/N_0 \log_2 M$ mayor que -1.6 dB, es posible la comunicación con cero error; la comunicación confiable no suele ser posible a menores razones de señal sobre ruido.

47.3.2 Codificación de la fuente

En la sección anterior se dijo que la información procedente de una fuente que produce distintos símbolos conforme a algún esquema de probabilidad puede definirse en términos de su entropía $H(X)$. Puesto que la entropía tiene unidades de bits por símbolo, también debe conocerse la velocidad de símbolos a fin de especificar la velocidad de información de la fuente en bits por segundo. En otras palabras, la velocidad de información de la fuente R_s está dada por:

$$R_s = rH(X) \text{ bits por segundo} \quad (47-33)$$

donde $H(X)$ es la entropía de la fuente en bits por símbolo y r es la velocidad de símbolos en símbolos por segundo.

Supóngase que esta fuente es la entrada a un canal con capacidad de C bits por símbolo o SC bits por segundo, donde S es la velocidad de símbolos disponi-

ble del canal. Un teorema importante de la teoría de la información, el teorema de la codificación sin ruido o primer teorema de Shannon, establece que, dados un canal y una fuente que genera información con menor velocidad que la capacidad del canal, es posible codificar la salida de la fuente de modo que pueda transmitirse por ese canal. Demostrar este teorema va más allá de los objetivos del presente análisis introductorio de la teoría de la información, pero el lector interesado puede consultar cualquiera de los textos estándares sobre teoría de la información. No obstante, aquí se demuestra de manera intuitiva ese teorema por medio de un ejemplo sencillo.

Ejemplo de codificación de la fuente

Considérese una fuente binaria discreta que tiene dos símbolos de salida posibles, A y B (fig. 47-31). Supóngase que la velocidad de símbolos r de la fuente es de 2.66 símbolos por segundo. Se supone que la salida de la fuente es la entrada a un canal que puede transmitir un 0 o un 1 binarios a una velocidad S de 1.5 símbolos por segundo con error despreciable. De esta forma, a partir del ejemplo 47.5 con $p = 1$, la capacidad del canal es de 1 bit por símbolo o:

$$SC = 1.5 \text{ bits por segundo}$$

Es claro que la velocidad de *símbolos* de la fuente excede la velocidad de *símbolos* del canal, de modo que los símbolos de la fuente no pueden ser directamente la entrada al canal.

Sin embargo, el primer teorema de Shannon indica que la velocidad de *información* de la fuente debe compararse con la capacidad del canal para determinar si es posible la transmisión por este último. Con objeto de realizar este cálculo se suponen dos conjuntos de probabilidades, que se ilustran en la tabla 47-1. Para el primer conjunto $P(A) = 0.9$ y $P(B) = 0.1$, y para el segundo $P(A) = 0.6$ y $P(B) = 0.4$. Se define un código mediante un mapeo de los símbolos de la fuente (A, B) en una sucesión de símbolos de código (0, 1). A

Tabla 47-1. Salidas de la fuente

Símbolos de la fuente	Código 1		Código 2	
	Probabilidad	Palabras de código	Probabilidad	Palabras de código
A	0.9	0	0.6	0
B	0.1	1	0.4	1
		$\bar{L} = 1$		$\bar{L} = 1$

Tabla 47-2. Salidas de la fuente extendida de segundo orden

Símbolos de la fuente	Código 1		Código 2	
	Probabilidad	Palabras de código	Probabilidad	Palabras de código
AA	0.81	0	0.36	00
AB	0.09	10	0.24	01
BA	0.09	110	0.24	10
BB	0.01	111	0.16	11
		$\bar{L} = 1.290$		$\bar{L} = 2$
		$\frac{1}{2} \bar{L} = 0.645$		$\frac{1}{2} \bar{L} = 1$

Tabla 47-3. Salidas de la fuente extendida de tercer orden

Símbolos de la fuente	Código 1		Código 2	
	Probabilidad	Palabras de código	Probabilidad	Palabras de código
AAA	0.729	0	0.216	00
AAB	0.081	100	0.144	010
ABA	0.081	101	0.144	011
BAA	0.081	110	0.144	100
ABB	0.009	11100	0.096	101
BAB	0.009	11101	0.096	110
BBA	0.009	11110	0.096	1110
BBB	0.001	11111	0.064	1111
		$\bar{L} = 1.598$		$\bar{L} = 2.944$
		$\frac{1}{3} \bar{L} = 0.533$		$\frac{1}{3} \bar{L} = 0.981$

fin de determinar si es posible la transmisión, es necesario calcular primero la entropía de la fuente. Para el primer conjunto de probabilidades la entropía es:

$$H_1(X) = -0.9 \log_2 0.9 - 0.1 \log_2 0.1$$

$$= 0.4690 \text{ bits por símbolo}$$

Para una velocidad de la fuente de 2.66 símbolos por segundo la velocidad de información es de:

$$rH_1(X) = 1.248 \text{ bits por segundo}$$

Esta cantidad es menor que la capacidad del canal de 1.5 bits por símbolo; así la transmisión es posible.

Para el segundo conjunto de probabilidades la entropía de la fuente es:

$$H_2(X) = -0.6 \log_2 0.6 - 0.4 \log_2 0.4$$

$$= 0.9710 \text{ bits por símbolo}$$

Esto produce una velocidad de información de:

$$rH_2(X) = 2.583 \text{ bits por segundo}$$

que excede la capacidad del canal, y en consecuencia la transmisión no es posible.

Sabiendo ya que con el primer conjunto de probabilidades es posible transmitir los símbolos de la fuente a través del canal, a continuación se busca un esquema de codificación adecuado.

Un esquema consiste en agrupar por pares los símbolos de la fuente y asignar la palabra de código más corta al par de símbolos de la fuente más probable, y la más larga al menos probable. Esto se muestra en la tabla 47-2. Los pares de símbolos (AA, AB, BA y BB) se conocen como salidas de la extensión de segundo orden de la fuente. La longitud promedio de palabra \bar{L} está dada por:

$$\bar{L} = \sum_{i=0}^{2^n} p_i l_i$$

donde p_i es la probabilidad de la i -ésima salida de la fuente extendida y l_i es la longitud de la i -ésima palabra de código. El parámetro n denota el orden de la extensión de la fuente, que en este caso es 2. La velocidad de símbolos a la salida del codificador es $r\bar{L}/n$. Así, para el código 1, la velocidad de símbolos a la salida del codificador es 1.716, que sigue excediendo la velocidad

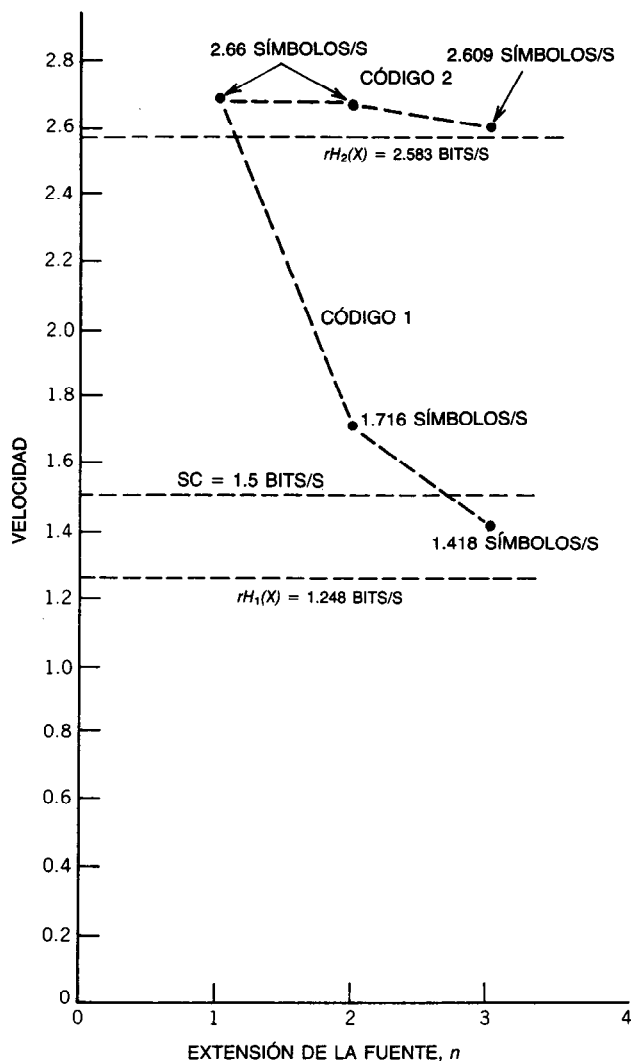


Fig. 47-32. Velocidades de información como función de la extensión de la fuente.

de símbolos permisible para el canal. Sin embargo, dicha velocidad se ha reducido a la salida del codificador, por lo que se sigue intentando.

En la tabla 47-3 se presenta la extensión de tercer orden de la fuente. Con este esquema \bar{L} es 1.598 y la velocidad de símbolos a la salida del codificador es $r\bar{L}/3 = 1.418$, de modo que el canal puede aceptarla.

En la figura 47-32 se ilustra el comportamiento de estos esquemas. Se observa que la velocidad de símbolos tiende a la velocidad de información $rH(X)$ a medida que n aumenta. Éste es un resultado fundamental de la teoría de la información. En la figura 47-32, los puntos representan velocidades de símbolos de la fuente en símbolos por segundo.

En este ejemplo las palabras de código fueron supuestas. A continuación se pasa al problema de determinar palabras de código.

Algunas definiciones

Antes de analizar el método para deducir palabras de código se presentan algunas definiciones útiles.

Las palabras de código se construyen con un *alfabeto*, que es una colección de símbolos empleados para la comunicación a través de un canal. Por ejemplo, una palabra de código binaria se construye con un alfabeto de dos símbolos, que suelen ser 0 y 1. La **longitud de palabra** de una palabra de código es el número de símbolos de esa palabra.

Existen algunas subdivisiones importantes de los códigos. Por ejemplo, un código puede ser *de bloques* o *de no bloques*. Un **código de bloques** es aquel en el que cada bloque de símbolos de la fuente se codifica en una sucesión con longitud fija de símbolos de código. Un código descifrable de forma única es un código de bloques en el cual las palabras de código pueden descifrar-

Tabla 47-4. Códigos instantáneo y no instantáneo

Símbolos de la fuente	Código 1 (no instantáneo)	Código 2 (instantáneo)
x_1	0	0
x_2	01	10
x_3	011	110
x_4	0111	1110

se sin necesidad de espacios. A su vez, estos códigos pueden clasificarse en instantáneos y no instantáneos, dependiendo de si es posible o no descodificar cada palabra en sucesión sin recurrir a símbolos de código posteriores. Alternativamente, en el caso de códigos no instantáneos es necesario recurrir a símbolos de código posteriores, como se ilustra en la tabla 47-4. Siempre debe recordarse que un código no instantáneo puede ser descifrable de forma única.

Una medida útil de la bondad de un código de fuente es la **eficiencia**, que se define como la razón de la longitud de palabra promedio mínima de las palabras de código, \bar{L}_{\min} , sobre la longitud de palabra promedio de la palabra de código, \bar{L} . Así:

$$\text{Eficiencia} = \frac{\bar{L}_{\min}}{\bar{L}} = \frac{\bar{L}_{\min}}{\sum_{i=1}^n p_i l_i} \quad (47-34)$$

donde, como antes, p_i es la probabilidad del i -ésimo símbolo de la fuente y l_i es la longitud de la palabra de código correspondiente al i -ésimo símbolo de la fuente.

Puede demostrarse que la longitud de palabra promedio mínima está dada por:

$$\bar{L}_{\min} = \frac{H(X)}{\log_2 D} \quad (47-35)$$

donde $H(X)$ es la entropía del mensaje total que se codifica y D es el número de símbolos en el alfabeto de codificación. De esta forma se obtiene:

$$\text{Eficiencia} = H(X)/\bar{L} \log_2 D \quad (47-36)$$

o bien:

$$\text{Eficiencia} = H(X)/\bar{L} \quad (47-37)$$

para un alfabeto binario, que es el caso usual.

Algunas veces se habla de **redundancia** de un código, que se define como:

$$\text{Redundancia} = 1 - \text{eficiencia} \quad (47-38)$$

Dado que el objetivo de la codificación de la fuente es elevar tanto como sea posible la eficiencia, se sigue que el objetivo final es eliminar la redundancia.

Codificación de Shannon-Fano

Existen varios métodos para codificar una salida de una fuente. Aquí se considerará sólo uno de ellos, la técnica de Shannon-Fano. Se elige para su estudio porque es fácil de aplicar y produce códigos razonablemente eficientes. Se estudiará por medio de un ejemplo.

Supóngase que se tiene un conjunto de salidas de una fuente por codificar. Primeramente se colocan estas salidas en orden no creciente de probabilidad de ocurrencia, como se ilustra en la tabla 47-5. El conjunto se divide entonces en dos subconjuntos (lo que se indica mediante la línea $A - A'$) equiprobables, y se asignan ceros al conjunto superior y unos al inferior, como se observa en la primera columna de palabras de código. Este proceso se repite, y cada vez los conjuntos se dividen de modo que tengan probabilidades tan semejantes como sea posible, hasta que ya no pueda hacerse otra partición. Este esquema producirá un código 100% eficiente si la partición siempre genera conjuntos equiprobables; de lo contrario el código será menos eficiente. Para este ejemplo particular:

$$\text{Eficiencia} = \frac{H(X)}{\bar{L}} = \frac{2.75}{2.75} = 1$$

dado que es posible la partición equiprobable.

Tabla 47-5. Codificación de Shannon-Fano

Palabras de la fuente	Probabilidad	Palabras de código	(Longitud)	(Probabilidad)
x_1	0.2500	00	2(0.25)	= 0.50
x_2	0.2500	01	2(0.5)	= 0.50
$A \cdots A'$				
x_3	0.1250	100	3(0.125)	= 0.375
x_4	0.1250	101	3(0.125)	= 0.375
x_5	0.0625	1100	4(0.0625)	= 0.25
x_6	0.0625	1101	4(0.0625)	= 0.25
x_7	0.0625	1110	4(0.0625)	= 0.25
x_8	0.0625	1111	4(0.0625)	= 0.25
Longitud de palabra promedio				= 2.75

Un procedimiento sugerido por Huffman produce un código con la menor longitud de palabra promedio, denominado código óptimo. Tal procedimiento, debido a que su aplicación resulta un poco más difícil que la del código de Shannon-Fano, no se abordará en esta sección.

47.3.3 Comunicaciones fiables con ruido

En seguida se comentan brevemente los métodos para lograr una comunicación confiable en presencia de ruido por medio del combate de los efectos de ese ruido. Cuando este estudio se emprende, los trabajos de Claude Shannon representan la promesa de considerable éxito.

El teorema de Shannon, algunas veces llamado teorema fundamental de la teoría de la información, establece que, dados un canal discreto sin memoria (cada símbolo es perturbado por ruido, independientemente de los demás símbolos) con capacidad C y una fuente con velocidad positiva R , donde $R < C$, existe un código tal que la salida de la fuente puede transmitirse por el canal con una probabilidad de error arbitrariamente pequeña.

De este modo, el teorema de Shannon predice la transmisión esencialmente libre de error en presencia de ruido. Por desgracia, este teorema sólo menciona la existencia de códigos y no especifica la forma de construirlos.

El teorema de Shannon demuestra la existencia de técnicas de codificación para la comunicación confiable. El desarrollo de técnicas prácticas de codificación ha dado por resultado avances significativos en la confiabilidad de las comunicaciones. Las más populares de estas técnicas se analizan más adelante, en la sección 48.5.

LIBROS Y REVISTAS

- N. Abramson, *Information Theory and Coding*, McGraw-Hill, Nueva York.
- A. B. Carlson, *Communication Systems*, McGraw-Hill, Nueva York.
- R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, Nueva York.
- B. P. Lathi, *Communication Systems*, Wiley, Nueva York.
- S. Lin, *An Introduction to Error-Correcting Codes*, Prentice-Hall, Englewood Cliffs, NJ.
- W. W. Peterson, *Error-Correcting Codes*, MIT, Cambridge, MA.
- F. M. Reza, *An Introduction to Information Theory*, McGraw-Hill, Nueva York.
- D. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information*, Wiley, Nueva York.
- M. Schwartz, *Information Transmission, Modulation, and Noise*, 2.^a ed., McGraw-Hill, Nueva York.
- C. E. Shannon, "A Mathematical Theory of Communications", *Bell Sys. Tech. J.* **27**:379-423, 623-656.
- C. E. Shannon y W. Weaver, *The Mathematical Theory of Communication*, University of Illinois, Urbana, IL.
- H. Taub y D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, Nueva York.
- N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, MIT, Cambridge, MA.
- P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Pergamon, Nueva York.

CAPÍTULO 48

Técnicas de comunicación

Jack W. Seyl

NASA Johnson Space Center, Houston, Texas

George W. Raffoul

Lockheed Engineering and Management Services
Company, Inc., Houston, Texas

Matthew J. Quinn, Jr.

College of Technology, University of Houston
Houston, Texas

K. K. Chow

Lockheed Missiles and Space Company, Palo Alto
Research Laboratory, Palo Alto, California

Bernard Sklar

Aerospace Corporation, Los Ángeles, California

Peter Monsen

P. M. Associates, Stow, Massachusetts

E. T. Dickerson

University of Houston-Clear Lake, Houston, Texas

48.1 Técnicas de modulación

- 48.1.1 Modulación lineal
- 48.1.2 Modulación y desmodulación angulares
- 48.1.3 Modulación de pulsos

48.2 Radiocomunicaciones

- 48.2.1 Introducción
- 48.2.2 Propagación de ondas de radio
- 48.2.3 Fundamentos sobre antenas
- 48.2.4 Fundamentos sobre transmisores
- 48.2.5 Fundamentos sobre receptores
- 48.2.6 Enlace de radiocomunicación total

48.3 Comunicaciones alámbricas

- 48.3.1 Definiciones
- 48.3.2 Modems para datos
- 48.3.3 Técnicas de modulación para modems
- 48.3.4 Velocidades de transmisión de datos
- 48.3.5 Tipos de servicio telefónico
- 48.3.6 Acondicionamiento de líneas
- 48.3.7 Conclusiones

48.4 Sistemas de comunicación por fibras ópticas

- 48.4.1 Primeros desarrollos de las fibras ópticas

- 48.4.2 Tipos de fibras y características de transmisión
- 48.4.3 Técnicas de fabricación de fibras
- 48.4.4 Componentes y sistemas de transmisión por fibras ópticas
- 48.4.5 Efectos de las radiaciones nucleares
- 48.4.6 Ejemplos sencillos de diseño
- 48.5 Técnicas avanzadas para las comunicaciones digitales**
 - 48.5.1 Introducción
 - 48.5.2 Etapas de procesamiento de la señal
 - 48.5.3 Análisis de enlaces de comunicaciones
- 48.6 Comunicaciones por canales con desvanecimiento**
 - 48.6.1 Introducción
 - 48.6.2 Canales de trayectoria múltiple con desvanecimiento lento
 - 48.6.3 Estructuras de receptores adaptativos
 - 48.6.4 Nuevas áreas de investigación
- 48.7 Análisis comparativo de las técnicas de comunicaciones**
 - 48.7.1 Criterios de comparación
 - 48.7.2 Comparación de técnicas de ondas de forma continua (analógica)
 - 48.7.3 Comparación de técnicas de ondas de forma discreta (digital)
 - 48.7.4 Elección entre técnicas de modulación digital y analógica
 - 48.7.5 Consideraciones prácticas
 - 48.7.6 Resumen

48.1 TÉCNICAS DE MODULACIÓN

Jack W. Seyl

La modulación puede definirse como el cambio en alguna característica de una señal u onda de forma conocida, que suele denominarse portadora, como función de alguna señal u onda de forma desconocida que transporta información.

En los sistemas de comunicación por radiofrecuencia (RF), en la generalidad de los casos la portadora es una onda de forma sinusoidal, y existen muchos métodos para modificar o modular la portadora. Algunos de tales métodos son:

- Modulación lineal.
- Modulación angular (o exponencial).
- Diversos tipos de modulación de pulsos.

En los siguientes párrafos se analizará con más detalle cada uno de ellos, pero primero se describen las características generales.

Una portadora sinusoidal es descrita por la función:

$$e_c(t) = E_c \sin(\omega_c t + \phi_0) \quad (48-1)$$

donde E_c = amplitud, V

ω_c = frecuencia, rad/s

ϕ_0 = fase inicial, rad

El argumento de la función seno de la ecuación 48-1 se denomina ángulo de fase, o ángulo de fase instantáneo, de la portadora. En la figura 48-1 se describen las características en el dominio del tiempo y en el dominio de la frecuencia de la portadora sinusoidal.

Como se observa en la expresión temporal sinusoidal de la ecuación 48-1, existen dos parámetros o características de la portadora que es posible modificar en correspondencia con una señal de información $g(t)$. Estos parámetros son la amplitud y el ángulo de fase.

Modulación lineal. En ella se hace variar la amplitud de la portadora de forma proporcional a la señal de información.

Por lo general se encuentran tres formas de modulación lineal:

- Modulación de amplitud (AM, de *amplitude modulation*).
- Modulación de doble banda lateral con supresión de la portadora (DSBSC, de *double side-band suppressed-carrier*).

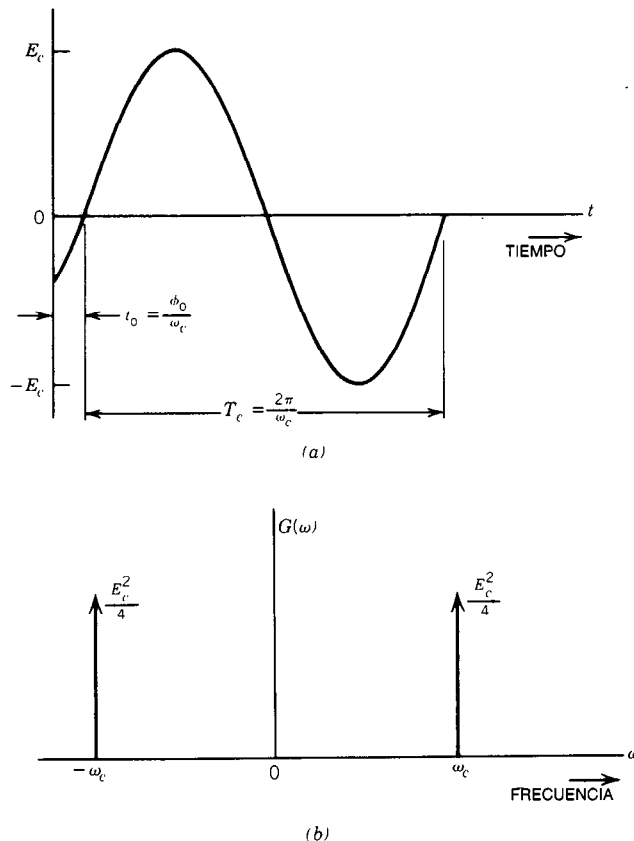


Fig. 48-1. Características de la señal portadora sinusoidal: a) función del tiempo; b) densidad espectral de potencia bilateral.

- Modulación de banda lateral única con supresión de la portadora (SSBSC, de *single sideband suppressed-carrier*).

En cada una de estas formas de modulación el proceso es lineal y equivalente a la multiplicación de la señal de información y la portadora. La portadora modulada resultante traslada efectivamente la energía de la señal de información a una banda de frecuencias (bandas laterales) inmediatamente adyacentes a la frecuencia de la portadora sinusoidal, permitiendo la transmisión por radiofrecuencia de la señal de información.

Modulación angular. También denominada modulación exponencial, en ella el argumento de la portadora sinusoidal se hace variar modulando ya sea la fase θ o la frecuencia ω . Cuando se hace variar la fase de manera directamente proporcional a la señal de información $g(t)$, entonces $\theta(t) = [\omega_c t + \phi(t) + \phi_0]$, lo que representa modulación de la fase (PM, de *phase modulation*). De forma alternativa, cuando la frecuencia ω de la portadora se hace variar en proporción directa con la señal de información, entonces $\theta(t) = [\omega(t)t + \phi_0]$, lo que representa modulación de frecuencia (FM, de *frequency modulation*).

Modulación de pulsos. Además de la modulación de ondas continuas, es posible utilizar una técnica para muestrear señales de información en el tiempo a fin de generar una señal modulada por pulsos (o de pulsos modulados). Si bien es posible utilizar directamente la señal de pulsos (pulsante) como la portadora de información, suele emplearse para modular una portadora de onda continua para efectos de transmisión. El muestreo en el tiempo también puede utilizarse para combinar o multicanalizar diversas señales de información para su transmisión en una sola portadora o canal de comunicación. Este proceso de muestrear y combinar en el tiempo se denomina multicanalización por división de tiempo (TDM, de *time division multiplexing*). Algunos de los parámetros pulsátiles que es posible modular o modificar por medio de una señal de información son la amplitud, duración y posición, y el resultado es:

- Modulación de pulsos en amplitud (PAM, de *pulse amplitude modulation*).
- Modulación de pulsos en duración (PDM, de *pulse duration modulation*).
- Modulación de pulsos en posición (PPM, de *pulse position modulation*).

Una forma más complicada de modulación de pulsos de amplio uso para los canales con gran capacidad de información es la modulación de pulsos en código o modulación por codificación de pulsos (PCM, de *pulse code modulation*). En los sistemas PCM, las señales individuales de información se muestrean en el tiempo y se cuantizan en amplitud. Después, la amplitud muestreada se transmite con una palabra binaria codificada que representa el intervalo de cuantización en

el que caen las muestras de amplitud. Cada palabra se transmite en un segmento de tiempo específico de un grupo de palabras, denominado cuadro. El instante de inicio del cuadro se identifica con una única palabra de sincronización.

La modulación en delta es una forma de cuantización predictiva, equivalente a un sistema de modulación de pulsos en código de 1 bit; es decir, cada palabra PCM se representa por medio de un solo bit. Tales sistemas se basan en la transmisión de las diferencias cuantizadas entre valores muestrales consecutivos, más que en las muestras mismas.

Además de la combinación de señales para la transmisión simultánea en tiempo, es costumbre separar las señales de información canalizando el dominio de la frecuencia, lo que da por resultado una señal multicanalizada por división de frecuencias (FDM, de *frequency-division multiplexed*). La señal FDM suele constar de varias subportadoras (sinusoides) que sirven para separar las diversas señales de información en el dominio de la frecuencia. La composición de las subportadoras se modula posteriormente en amplitud, fase o frecuencia sobre una portadora de radiofrecuencia para su transmisión.

En años recientes ha aparecido otra forma de canal de comunicación de acceso múltiple. En estos canales se utilizan códigos pseudoaleatorios para identificar a cada usuario y para separar la señal de un usuario de la de otro cuando se reciben en un receptor común. Tales técnicas también proporcionan algún grado de seguridad y protección a la propagación por trayectoria múltiple.

48.1.1 Modulación lineal

La modulación lineal de una portadora sinusoidal se obtiene cuando se modifica su amplitud instantánea como función lineal de la señal de información. Las señales moduladas linealmente pueden transmitirse en varios formatos. En la generalidad de los casos, las señales se clasifican conforme a las características de sus espectros de potencia. En el caso de la modulación en amplitud (AM), el espectro de potencia contiene una componente portadora con energía espectral de la señal de información por encima y por debajo de la frecuencia de la portadora. En la modulación de doble banda lateral con supresión de la portadora (DSBSC), la señal de información tiene energía espectral a frecuencias por encima y por debajo de la frecuencia de la portadora; sin embargo, la componente de frecuencia de la portadora se suprime en el espectro de potencia de la señal modulada. En la modulación de banda lateral única con supresión de la portadora (SSBSC), el espectro de la señal de información se presenta ya sea por encima o por debajo de la frecuencia de la portadora (banda lateral única), y se suprime la componente de frecuencia de la portadora.

Independientemente del formato de modulación deseado, la modulación lineal puede considerarse un proceso de multiplicación en el cual la función de tiempo de la señal de información se multiplica por la función de tiempo de la señal portadora. Este proceso da

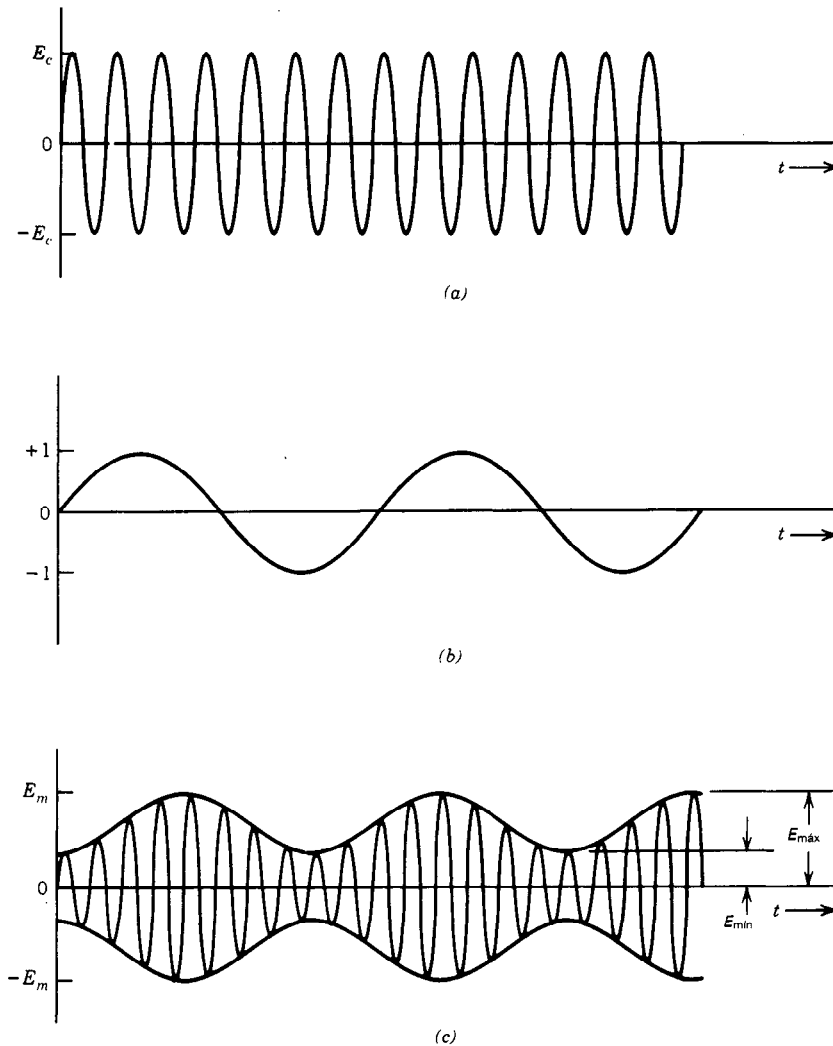


Fig. 48-2. Funciones del tiempo de la modulación de amplitud: a) señal portadora, $e_c(t) = E_c \sin \omega_c t$ (suponiendo que $\phi_0 = 0$); b) señal de información, $g(t) = \sin \omega_m t$ (suponiendo que $\phi_m = 0$); c) señal modulada en amplitud, $e_{AM}(t) = E_c[1 + m \sin \omega_m t] \sin \omega_c t$.

por resultado la traslación simétrica del espectro de la señal de información con respecto a la frecuencia de la portadora.

Modulación en amplitud

En este tipo de modulación, la amplitud instantánea de la portadora (sinusoide) se hace variar en forma proporcional a la amplitud de la señal moduladora. En el caso general, la descripción en el dominio del tiempo de una señal AM es:

$$e_{AM}(t) = [1 + mg(t)]E_c \sin(\omega_c t + \phi_0) \quad (48-2)$$

donde $g(t)$ = señal de información

E_c = amplitud pico de la portadora sin modular, V

ω_c = frecuencia de la portadora, rad/s

ϕ_0 = ángulo de fase inicial de la portadora, rad

m = índice de modulación

En la figura 48-2 se muestra la función instantánea de tiempo para el caso especial en que $g(t)$ es una señal de información sinusoidal con amplitud normalizada a la unidad:

$$g(t) = \sin(\omega_m t + \phi_m)$$

donde ω_m es la frecuencia de la señal de información en radianes por segundo y ϕ_m es la fase de la señal de información en radianes.

El índice de modulación de la señal AM determina la cantidad de energía contenida en las bandas laterales

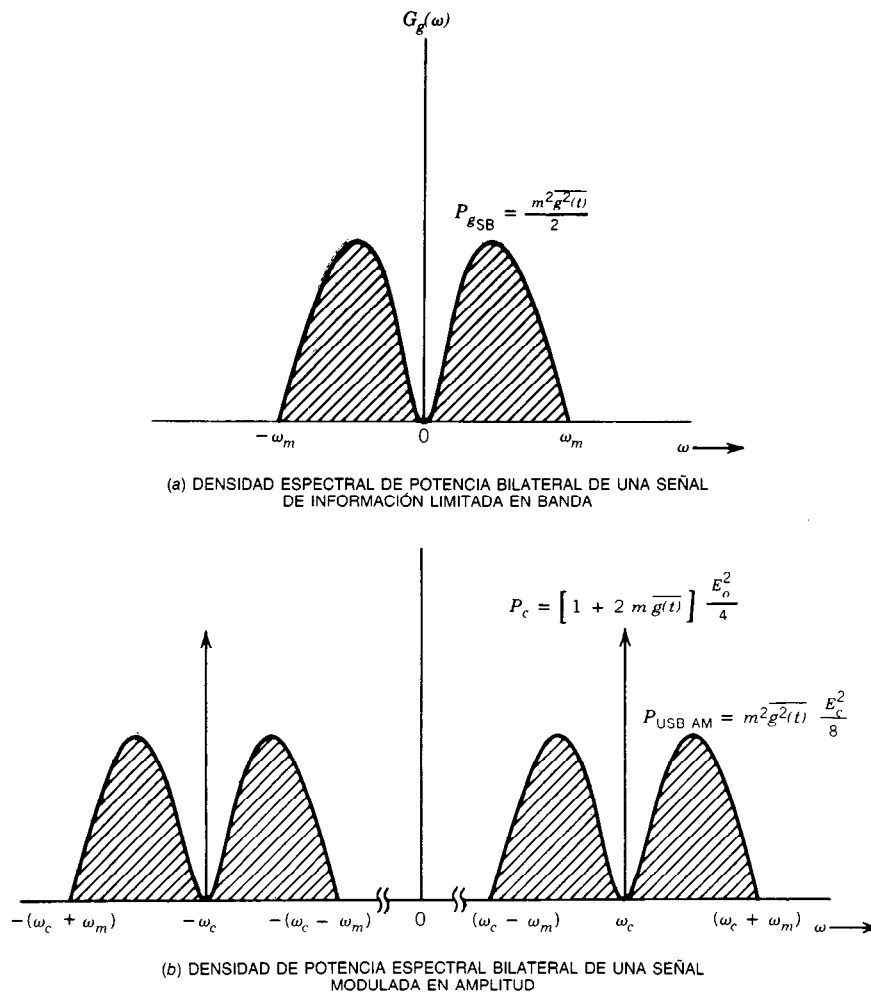


Fig. 48-3. Densidad espectral de potencia, modulación en amplitud (AM) con señal de información limitada en banda. a) Densidad espectral de potencia bilateral de la señal de información limitada en banda. b) Densidad espectral de potencia bilateral de la señal modulada en amplitud.

de la señal. En la señal AM $e_{AM}(t)$ de la figura 48-2, el índice de modulación m está definido como:

$$m \% = \left(\frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \right) \times 100 \quad (48-3)$$

donde E_{\max} es el máximo (cresta) de la amplitud de la señal modulada y E_{\min} es el mínimo (valle) de la amplitud de la señal modulada.

En la mayor parte de los sistemas AM, el índice está restringido a no más del 100% ($m = 1.0$), para evitar distorsión no lineal de la señal de información y la generación de armónicos de la portadora transmitida.

Densidad espectral de potencia y potencia de la banda lateral. La distribución de la potencia transmitida en una señal AM se describe por medio de la función densidad espectral de potencia, que es:

$$G_{AM}(\omega) = [1 + 2m \overline{g(t)}] \frac{E_c^2}{4} [\delta(\omega - \omega_c) + \delta(\omega + \omega_c)] + \frac{E_c^2}{4} [G_g(\omega - \omega_c) + G_g(\omega + \omega_c)] \quad (48-4)$$

donde $G_g(\omega)$ es el espectro de potencia de la señal de información $g(t)$, y $\delta(\omega)$ representa la función delta de Dirac.

Si la señal de información está limitada en banda a ω_m rad/s, el espectro de potencia de la señal AM está limitado en banda a la frecuencia de la portadora $\pm \omega_m$.

La potencia total en la señal AM se expresa de la siguiente forma:

$$P_{AM} = [1 + 2 m \overline{g(t)} + m^2 \overline{g^2(t)}] E_c^2/2 \quad (48-5)$$

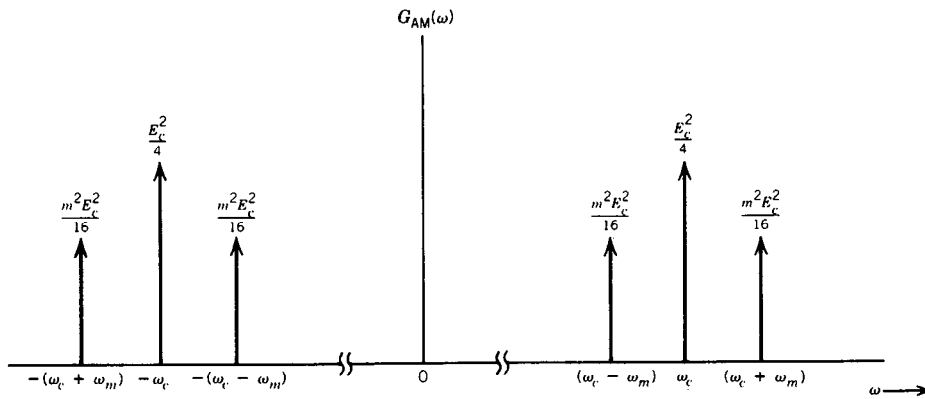


Fig. 48-4. Espectro de potencia de una señal modulada en amplitud con una señal de modulación (de tono) sinusoidal [$e_m(t) = \text{sen } \omega_m t$].

donde m = índice de modulación

$\overline{g(t)}$ = valor medio de $g(t)$

$\overline{g^2(t)}$ = valor cuadrático medio de $g(t)$

En la figura 48-3 se muestran las funciones densidad espectral de potencia de la señal de información y de la señal AM.

La potencia total en las bandas laterales está dada por:

$$P_{SB} = [m^2 \overline{g^2(t)}] \frac{E_c^2}{2} \quad (48-6)$$

Para el caso especial en que la señal de modulación es una senoide con amplitud unitaria, en la figura 48-4 se muestra la densidad espectral de potencia bilateral. Dado que $g(t) = \text{sen}(\omega_c t + \phi_0)$, el valor medio de $g(t)$ es cero, y el valor cuadrático medio es igual a $1/2$. Por tanto:

$$P_{AM} = \frac{E_c^2}{2} \left(1 + \frac{m^2}{2} \right) \quad (48-7)$$

y

$$P_{SB} = m^2 \frac{E_c^2}{4} \quad (48-8)$$

Como se observa con referencia a las ecuaciones para el caso especial de una señal de información sinusoidal, y suponiendo el máximo índice de modulación del 100% ($m = 1.0$), sólo la tercera parte de la potencia total está contenida en las bandas laterales de información.

Modulación de doble banda lateral con supresión de la portadora (DSBSC)

Si la señal de modulación no contiene término constante, el formato de modulación cambia al de la doble banda lateral con supresión de la portadora (DSBSC). La representación en el dominio del tiempo de la señal DSBSC es:

$$e_{DSB}(t) = mg(t)\text{sen}(\omega_c t + \phi_0) \quad (48-9)$$

donde m , $g(t)$, ω_c y ϕ_0 significan lo mismo que en el caso de AM.

La estructura en el dominio del tiempo de la onda de doble banda lateral con supresión de la portadora, suponiendo que la señal de modulación $g(t)$ sea una onda sinusoidal, se muestra en la figura 48-5. Cada vez que la señal de información pasa por el valor cero, la fase de la señal portadora se desplaza 180° (es decir, existe una inversión de fase). Esta característica de la onda DSBSC es importante en los sistemas de transmisión digital, ya que permite la generación de una técnica digital especial de modulación de fase, denominada manipulación por desplazamiento de fase (PSK, de *phase shift keying*). Las características de la modulación PSK o DSBSC digital se describen en detalle en la sección dedicada a la modulación digital de fase.

Densidad espectral de potencia y potencia de banda lateral

La densidad espectral de potencia de una señal DSBSC es:

$$G_{DSB}(\omega) = \frac{E_c^2}{4} [G_g(\omega - \omega_c) + G_g(\omega + \omega_c)] \quad (48-10)$$

En la figura 48-6 se ilustra la densidad espectral de potencia bilateral de la señal de información de banda base limitada, así como de la onda modulada DSBSC.

Es de interés observar que el efecto de la modulación DSBSC es una traslación del espectro de modulación de la banda base a la frecuencia de la portadora. También se observa que el ancho de banda ocupado por la energía espectral de la señal de información es dos veces el del espectro de la banda base:

$$BW_{DSB} = [(\omega_c + \omega_m) - (\omega_c - \omega_m)] = 2\omega_m \quad (48-11)$$

Esta duplicación efectiva del ancho de la banda base también ocurría para el espectro de la señal AM. La potencia total de la señal DSBSC es:

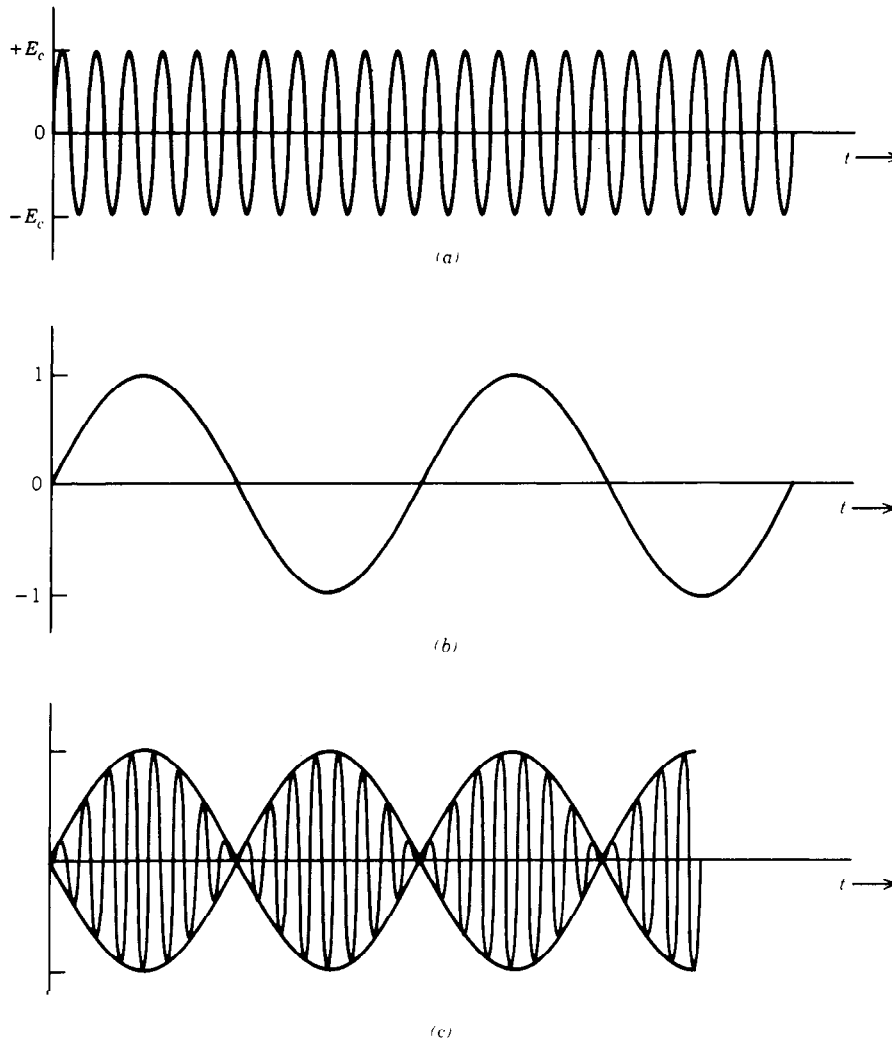


Fig. 48-5. Funciones del tiempo de la modulación de doble banda lateral con supresión de la portadora (DSBSC): a) señal portadora RF, $e_c(t) = E_c \sin(\omega_c t + \phi_0)$; b) señal de información, $g(t) = \sin \omega_m t$ (suponiendo que $\phi_m = 0$); c) señal modulada DSBSC $e_{DSB}(t) = E_c \sin \omega_m t \sin(\omega_c t + \phi_0)$.

$$P_{DSB} = \frac{E_c^2}{2} \overline{g^2(t)} \quad (48-12)$$

donde $g^2(t)$ es el valor cuadrático o potencia total de la señal de información.

La potencia total de la banda lateral de la DSBSC es:

$$P_{SB,(DSB)} = \frac{E_c^2}{2} \overline{g^2(t)} \quad (48-13)$$

Por tanto, toda la potencia de una onda DSBSC está contenida en el espectro de la señal de información, lo que da por resultado un proceso de transmisión más eficiente que el de AM. Sin embargo, el proceso

de desmodulación para las señales moduladas con supresión de la portadora es considerablemente más complicado que para el caso de la AM, y así existe una consideración de complejidad y costo del receptor en comparación con la eficiencia del transmisor.

Modulación de banda lateral única con supresión de la portadora (SSBSC)

Una señal de banda lateral única con supresión de la portadora puede generarse mediante técnicas de filtrado (hacer pasar sólo una de las bandas laterales de modulación) o de defasamiento durante el proceso de modulación. Si se hace pasar una onda de doble banda lateral con supresión de la portadora a través de un filtro que elimine por completo todas las señales por

debajo de la portadora, como se muestra en la figura 48-7, se obtiene una onda de banda lateral única con supresión de la portadora.

Es posible obtener un resultado parecido mediante el empleo de la denominada técnica de discriminación de fase, que requiere la adición de una versión de cuadratura de la señal modulada de doble banda lateral con supresión de la portadora a la onda original DSBSC (Panter). En la figura 48-8 se ilustra esta técnica de modulación SSBSC.

La representación resultante de la función temporal para la modulación SSBSC puede expresarse con mayor brevedad en forma de cuadratura de la portadora; es decir:

$$e_{SSB}(t) = g(t) \cos \omega_c t - U(t) \sin \omega_c t$$

donde $U(t)$ es la versión de cuadratura de $g(t)$ obtenida al introducir un defasamiento constante de 90° en todas las frecuencias de $g(t)$.

Densidad espectral de potencia y potencia de la banda lateral

El espectro de potencia para la onda SSBSC es esencialmente el de la señal DSBSC sin una de las bandas laterales de información. En la figura 48-9 se ilustra el espectro SSBSC cuando se ha elegido la banda lateral para transmisión.

El espectro de densidad de potencia está dado por la expresión:

$$G_{SSB}(\omega) = \frac{E_c^2}{4} [G_g(\omega - \omega_c) + G_g(\omega + \omega_c)] \text{ para } |\omega| > \omega_c$$

$$0 \text{ para } |\omega| < \omega_c \quad (48-14)$$

La potencia total de la onda SSBSC es:

$$P_{SSB} = \frac{E_c^2}{4} \overline{g^2(t)} \quad (48-15)$$

y la potencia de la banda lateral es:

$$P_{SB} = \frac{E_c^2}{4} \overline{g^2(t)} \quad (48-16)$$

Como en el caso de la señal DSBSC, la potencia total y la potencia de la banda lateral de la señal de información son iguales para la señal SSBSC. Sin embargo, la potencia total disponible es la mitad que en el caso de la señal modulada DSBSC si la amplitud de la portadora, E_c , es igual.

Detección de la modulación lineal

Los espectros transmitidos de las diferentes formas de señales moduladas linealmente varían en magnitud en

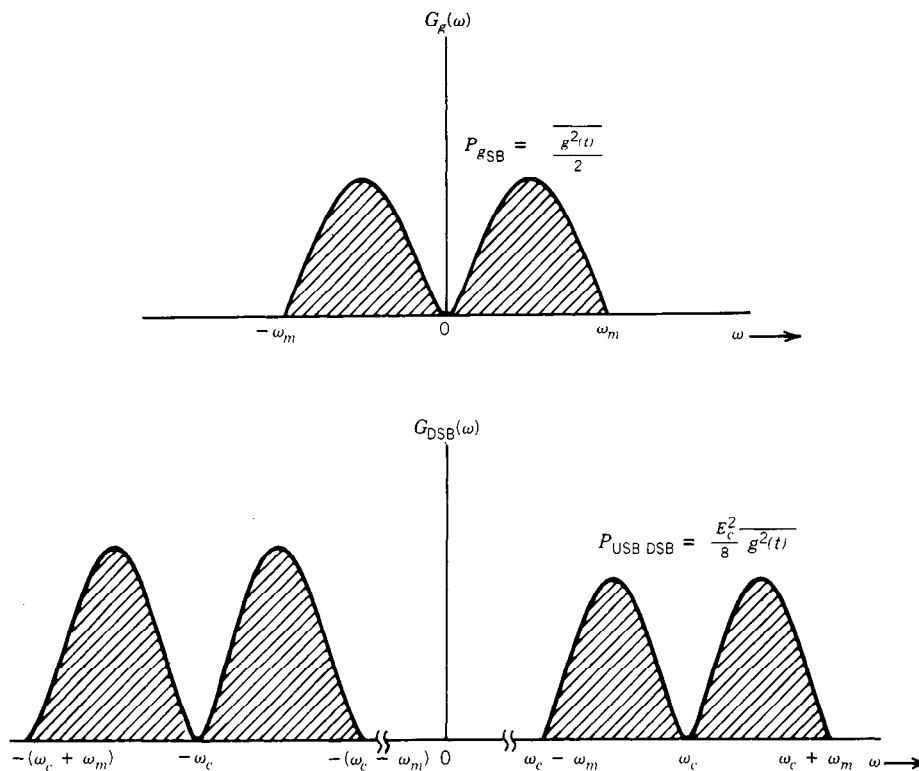


Fig. 48-6. Espectros de la densidad espectral de potencia: a) de una señal de información limitada en banda; b) de una señal de doble banda lateral con supresión de la portadora.

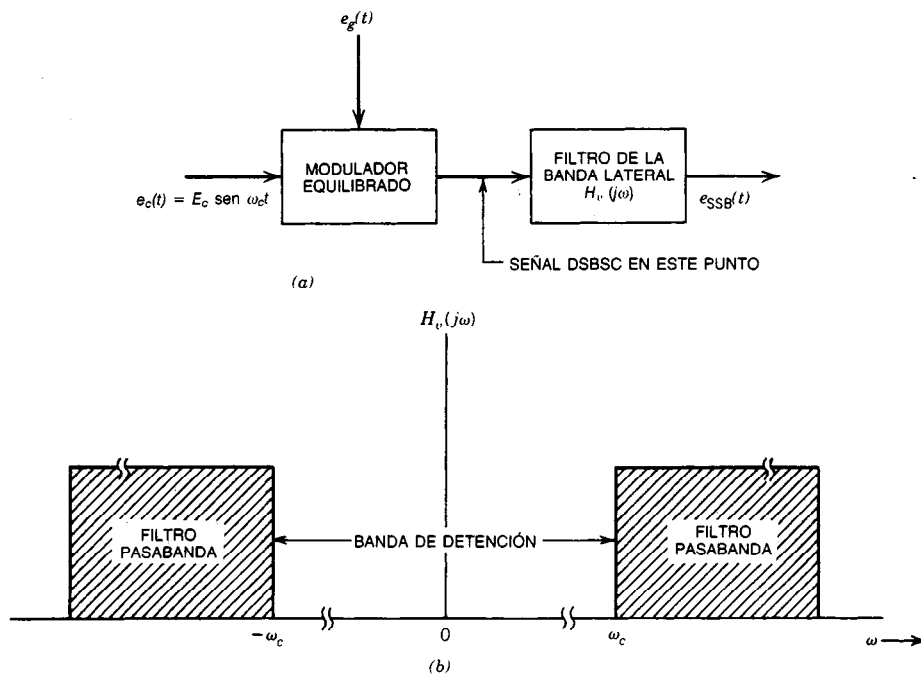


Fig. 48-7. Generación de una señal de banda lateral única con supresión de la portadora (SSBSC), utilizando un modulador de doble banda lateral con supresión de la portadora (DSBSC) y filtrado de la banda lateral: *a*) diagrama funcional del modulador SSBSC; *b*) características del filtro de la banda lateral (se muestra para el caso de la banda lateral superior).

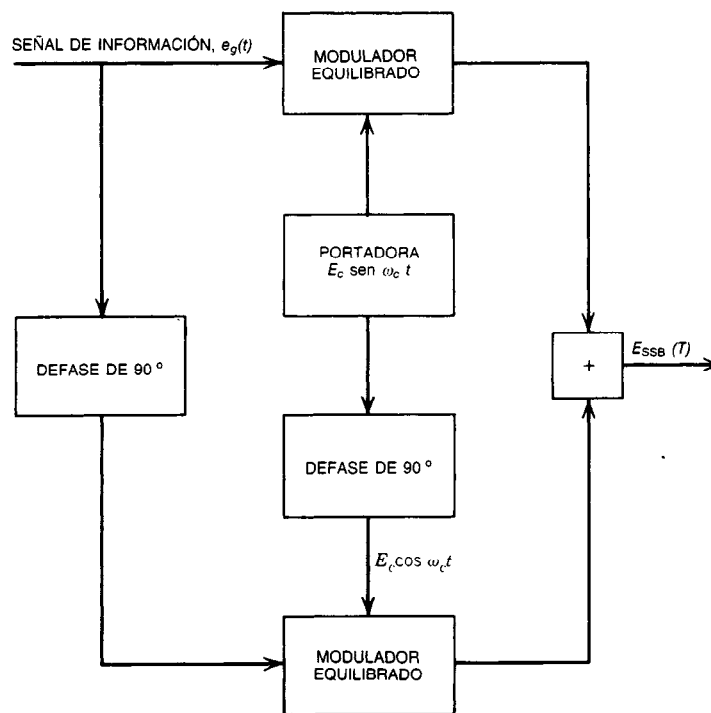


Fig. 48-8. Técnica de desplazamiento de fase (defasamiento) para la generación de la modulación de banda lateral única con supresión de la portadora (SSBSC).

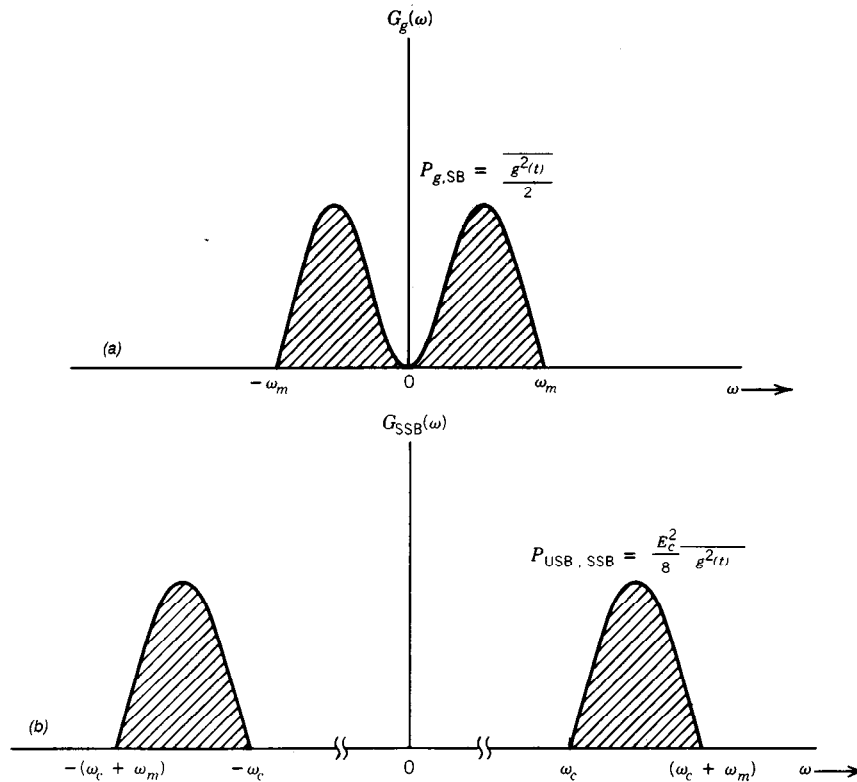


Fig. 48-9. Densidad espectral de potencia para a) el espectro de la banda base de la señal de información bilateral, suponiendo limitación en banda a $\pm\omega_m$, y b) espectro de una onda modulada de banda lateral única con supresión de la portadora bilateral (SSBSC).

Tabla 48-1. Comparación de las técnicas de modulación y desmodulación lineal (restringidas en potencia de banda lateral)

	Tipo de modulación		
	Modulación de amplitud (AM) ^a	De doble banda lateral con supresión de la portadora (DSBSC)	De banda lateral única con supresión de la portadora (SSBSC)
Potencia de la banda lateral (W)	$\overline{g^2(t)}P_c^b$	$\overline{g^2(t)}P_c$	$\overline{g^2(t)}P_c/2$
Ancho de la banda de predetección bilateral (Hz)	$4\omega_m$	$4\omega_m$	$2\omega_m$
Densidad espectral de la potencia del ruido (W/rad/s)	$N_0/2$	$N_0/2$	$N_0/2$
Señal sobre ruido de predetección $(S/N)_i$	$\frac{\overline{g^2(t)}P_c}{2N_0\omega_m}$	$\frac{\overline{g^2(t)}P_c}{2N_0\omega_m}$	$\frac{\overline{g^2(t)}P_c}{2N_0\omega_m}$
Señal sobre ruido de salida $(S/N)_0^c$	$\frac{\overline{g^2(t)}P_c}{P_n}$	$\frac{\overline{g^2(t)}P_c}{P_n}$	$\frac{\overline{g^2(t)}P_c}{P_n}$

^a Índice de modulación = 1.0.

^b $P_c = \alpha E_c^2/2$ = potencia portadora disponible en el detector; α es una constante de atenuación.

^c $P_n = 2N_0\omega_m$.

Tabla 48-2. Comparación de las técnicas de modulación y desmodulación lineal (normalizadas a la potencia total de la señal)

	Tipo de modulación		
	Modulación de amplitud (AM) ^a	De doble banda lateral con supresión de la portadora (DSBSC)	De banda lateral única con supresión de la portadora (SSBSC)
Potencia total de la señal P_T , W	$3P_c/2$	$P_c/2$	$P_c/4$
Potencia de la banda lateral ^c P_{SB}	$P_c/2 = P_T/3$	$P_c/2 = P_T$	$P_c/4 = P_T$
Señal sobre ruido de predetección $(S/N)_i$	$P_T/6N_0\omega_m$	$P_T/2N_0\omega_m$	$P_T/N_0\omega_m$
Señal sobre ruido de salida $(S/N)_o$ ^d	$P_T/3P_n$	P_T/P_n	$2P_T/P_n$

^a Modulación de onda sinusoidal.

^b $m = 1.0$.

^c $P_c = aE_c^2/2$.

^d $P_n = 2N_0\omega_m$.

distribución de frecuencia. La recuperación de información de la señal portadora recibida requiere un proceso de conversión y detección de frecuencia.

En la detección de doble banda lateral y banda lateral única con supresión de la portadora, ésta debe ser restaurada en el receptor antes de que pueda ocurrir la desmodulación. Esta forma de detección suele denominarse detección coherente o sincrónica (síncrona), ya que es necesario obtener en el receptor una señal portadora en fase coherente o sincrónica con la señal portadora transmitida. Para las señales AM ordinarias la detección sincrónica es innecesaria, ya que la señal de información está contenida por completo en la envolvente de la portadora y no en la fase.

Esto da por resultado una clara ventaja en complejidad y costo de los receptores ordinarios AM con respecto a los receptores DSBSC y SSBSC, ya que el sistema de detección puede ser sustancialmente más sencillo si no se requiere reconstrucción de la portadora.

El rendimiento de las técnicas de modulación y desmodulación lineal puede compararse suponiendo que la potencia disponible de la portadora antes de la modulación, $P_c = E_c^2/2$, es constante. En la tabla 48-1 se resumen los resultados y se muestra la potencia de la banda lateral en términos de la potencia disponible de la portadora. En el caso de la SSBSC, la potencia de la banda lateral es la mitad de la propia de la AM y la DSBSC, como ya se analizó.

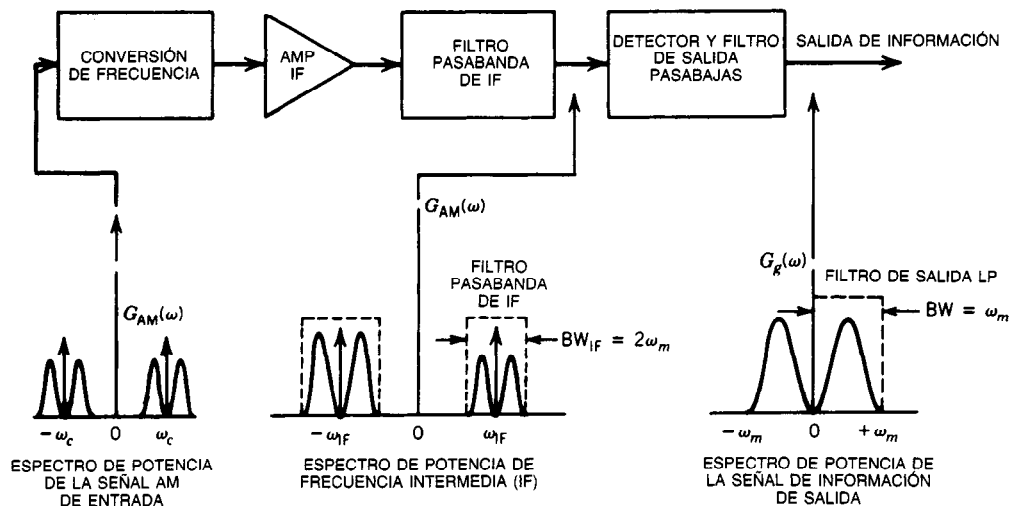


Fig. 48-10. Recepción y detección por AM ordinarias.

Utilizando el modelo de doble banda en el dominio de la frecuencia, el ancho de banda requerido del receptor antes de la detección es de $4\omega_m$ para AM y DSBSC, pero sólo $2\omega_m$ para SSBSC. Si en la práctica el receptor se encuentra restringido a este ancho de banda, la relación de señal sobre ruido de la predetección, $(S/N)_i$, es exactamente la misma para SSBSC que para DSBSC y AM. De este modo el rendimiento global, medido por la relación de señal sobre ruido de salida $(S/N)_o$, es igual para los tres casos.

En los sistemas SSBSC prácticos es costumbre incrementar la salida de potencia pico, de modo que la potencia media de salida (potencia de la banda lateral) sea la misma que sería en el caso DSBSC. Así, la relación señal sobre ruido de salida para la SSBSC es dos veces la de la AM o de la DSBSC.

En la tabla 48-2 se muestra una comparación parecida, pero con la relación señal sobre ruido de salida normalizada a la potencia total de la señal recibida, P_T . Dado que la AM requiere la transmisión de una portadora, solamente la tercera parte de la potencia total de la señal se encuentra en las bandas laterales, por lo que el rendimiento global para una potencia de transmisión fija es más deficiente que para la DSBSC o para la SSBSC.

En esta tabla, la SSBSC parece tener una ventaja de 2:1 (3 dB) sobre la DSBSC, debido a que la SSBSC solamente requiere la mitad del ancho de banda de predetección. Esta ventaja se cumple si la potencia total recibida es igual en ambos casos. Si las potencias totales de la banda lateral son iguales, entonces se cumplen los resultados de la tabla 48-1.

En las figuras 48-10, 48-11 y 48-12 se ilustra el proceso de recepción/detección para AM, DSBSC y SSBSC, respectivamente.

48.1.2 Modulación y desmodulación angulares

Como ya se observó, un parámetro de la portadora que es posible modular o variar a fin de transportar información es el ángulo de fase.

Si en el argumento de la función sinusoidal se incluye un término que varía proporcionalmente con la señal de modulación o de información, el resultado es la modulación de fase (PM, de *phase modulation*).

Por tanto, para la PM la ecuación en el dominio del tiempo es:

$$e_{PM}(t) = E_c \sin [\omega_c t + \theta(t) + \phi_0] \quad (48-17)$$

donde $\theta(t) = \Delta\theta g(t)$

$\Delta\theta$ = índice de modulación de fase, rad

$g(t)$ = función temporal de la señal de información

ω_c = frecuencia de la portadora, rad/s

ϕ_0 = ángulo de fase inicial arbitrario, rad

Si el argumento es tal que la diferencia en la frecuencia instantánea $\omega_i(t)$ y la frecuencia de la portadora es proporcional a la señal de modulación o de información, entonces la señal es modulada en frecuencia (FM, de *frequency modulated*). Así, para la FM la frecuencia varía proporcionalmente a la señal de información $g(t)$:

$$\omega_i(t) - \omega_c = \Delta\omega g(t)$$

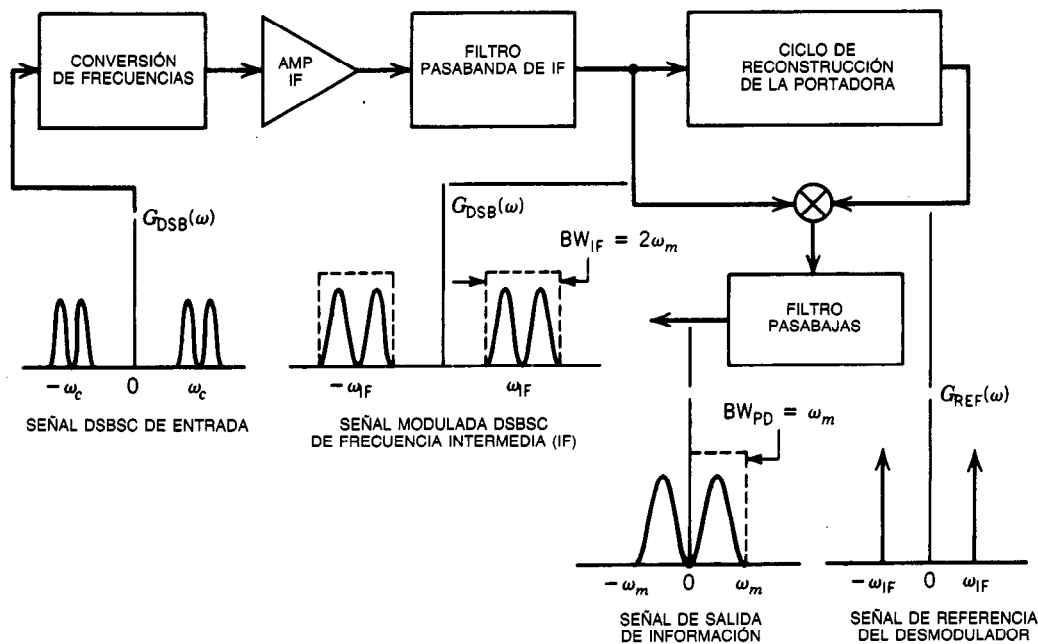


Fig. 48-11. Recepción y detección de doble banda lateral con supresión de la portadora.

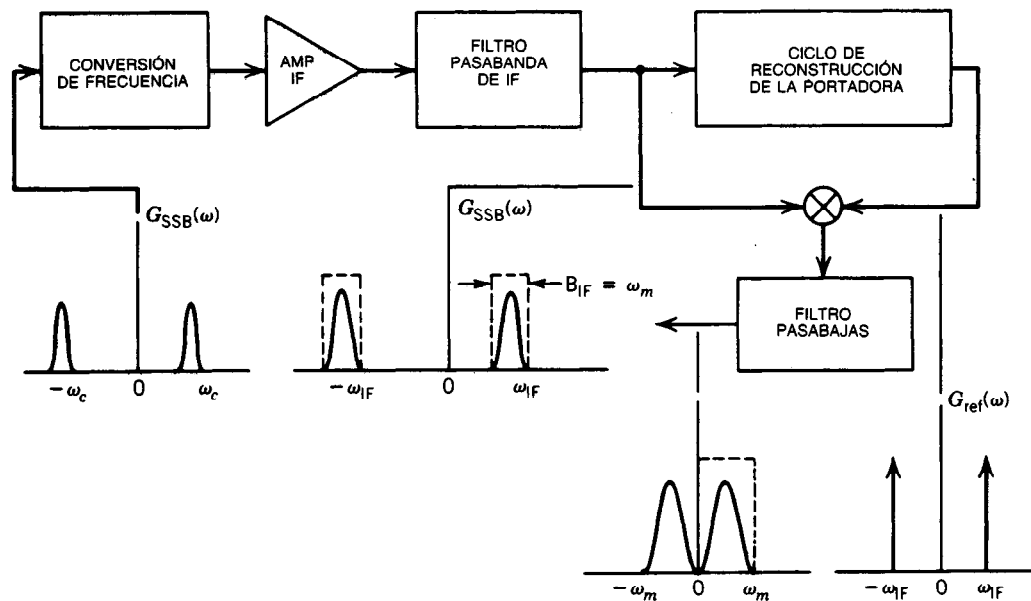


Fig. 48-12. Recepción y detección de banda lateral única con supresión de la portadora.

Ya que $\omega_i(t) = d\theta(t)/dt$, para la FM la ecuación de la onda de forma en el dominio del tiempo es:

$$e_{FM}(t) = E_c \sin[\omega_c t + \theta(t) + \phi_0] \quad (48-18)$$

donde $\theta(t) = \Delta\omega \int_{-\infty}^t g(t)dt$.

$\Delta\omega$ es la desviación de la frecuencia de cresta en radianes por segundo. Por consiguiente, en general el senoide modulado en ángulo puede expresarse en forma compleja como el fasor:

$$\vec{E}_\theta(t) = E_c \exp[j(\omega_c t + \phi_0)] \exp[j\theta(t)] \quad (48-19)$$

donde $j = \sqrt{-1}$. La expresión fasorial revela que, para el caso de la modulación en ángulo, la portadora $E_c \exp[j(\omega_c t + \phi_0)]$ ha sido modificada por una función exponencial del mensaje $\theta(t)$. Por ello, los términos "modulación exponencial" y "modulación angular" son sinónimos. Como se observa en la estructura temporal de la portadora modulada en ángulo, la amplitud de la señal permanece constante cuando se aplica la modulación. En la figura 48-13 se ilustra una portadora modulada en ángulo cuando la señal de información es una sola onda sinusoidal $g(t) = \Delta\theta \sin \omega_m t$. Las señales resultantes PM y FM difieren en que la variación de fase para el caso PM es una función sinusoidal, y para FM es una función cosenoidal.

A fin de desmodular o recuperar la señal de información, la portadora modulada en ángulo debe ser procesada por el sistema receptor para obtener una señal de salida proporcional a su fase o frecuencia instantáneas. Para el caso PM, la detección puede efectuarse empleando desmodulaciones de productos, donde la señal de referencia multiplicadora se encuentre en cuadratura de fase con la portadora transmitida. Para la detección de FM se requiere el empleo de un discriminador de frecuencia, que en su forma más sencilla es una red que primero convierte la modulación de frecuencia a modulación de amplitud y después realiza la detección de la envolvente de la señal AM.

Modulación de fase

Debido a las características no lineales del proceso PM, el espectro de potencia de la portadora modulada resultante es más complicado que en la modulación lineal. Por consiguiente, en el análisis del espectro de potencia de la onda PM solamente se considerarán unos cuantos tipos de señales de modulación.

Cuando $g(t) = \sin \omega_m t$, para la onda modulada en fase la función en el dominio del tiempo es:

$$e_{PM}(t) = E_c \sin[\omega_c t + \Delta\theta \sin \omega_m t] \quad (48-20)$$

que puede desarrollarse (Giacoletto) como:

$$e_{PM}(t) = E_c [J_0(\Delta\theta) \sin \omega_c t + \sum_{n=1}^{\infty} J_{2n}(\Delta\theta) [\sin(\omega_c + 2n\omega_m)t + \sin(\omega_c - 2n\omega_m)t] + \sum_{n=1}^{\infty} J_{2n-1}(\Delta\theta) \{\sin[\omega_c + (2n-1)\omega_m]t - \sin[\omega_c - (2n-1)\omega_m]t\}] \quad (48-21)$$

donde $J_x(\Delta\theta)$ es la función de Bessel de primera clase y orden x , con argumento igual al índice de modulación en fase $\Delta\theta$.

Así, el espectro de potencia de una onda PM modulada sinusoidalmente es una serie de funciones de pulso espaciadas simétricamente con respecto a la frecuencia portadora ω_c en múltiplos de la frecuencia de modulación ω_m .

El peso de cada pulso es determinado por la función de Bessel de orden idóneo para la banda lateral específica en cuestión.

Teóricamente, en el espectro PM aparece una infinidad de bandas laterales. Sin embargo, las características de la función de Bessel son tales que las magnitudes de la banda lateral decrecen con rapidez a medida que aumenta el orden. En la figura 48-14 se muestran los coeficientes de Bessel para las primeras bandas laterales de noveno orden de una onda PM o FM como función de su argumento $\Delta\theta$ o β .

La modulación de fase extrae potencia de la señal portadora y la deposita en las bandas laterales. Para la modulación de tono único, el espectro de potencia es como se ilustra en la figura 48-15, suponiendo un ín-

dice de modulación de fase $\Delta\theta$ constante y una frecuencia ω_m creciente de la señal de modulación.

La potencia en las bandas laterales y las componentes de la portadora de la señal PM es (Painter y Hondros):

$$P_{PM} = \frac{E_c^2}{2} \left[\sum_{-\infty}^{\infty} J_n^2(\Delta\theta) \right] \quad (48-22)$$

Dado que $\sum_{-\infty}^{\infty} J_n^2(\Delta\theta) = 1$, la potencia total de la onda modulada es equivalente a la propia de la portadora sin modular $E_c^2/2$. Si se utiliza más de una señal de onda sinusoidal para modular la fase de la misma portadora, el espectro resultante será una interrelación compleja de las frecuencias y los índices de modulación de las dos señales. De hecho, si las dos señales de modulación no están relacionadas coherente o armónicamente, el espectro de potencia consistirá en pulsos a frecuencias cercanas a la propia de la portadora equivalentes a la suma y las diferencias de cada una de las señales de modulación y sus frecuencias armónicas. En el caso general de la modulación PM por medio de K ondas sinusoidales, la expresión para la onda PM en el dominio del tiempo pasa a ser (Giacoletto, 1947):

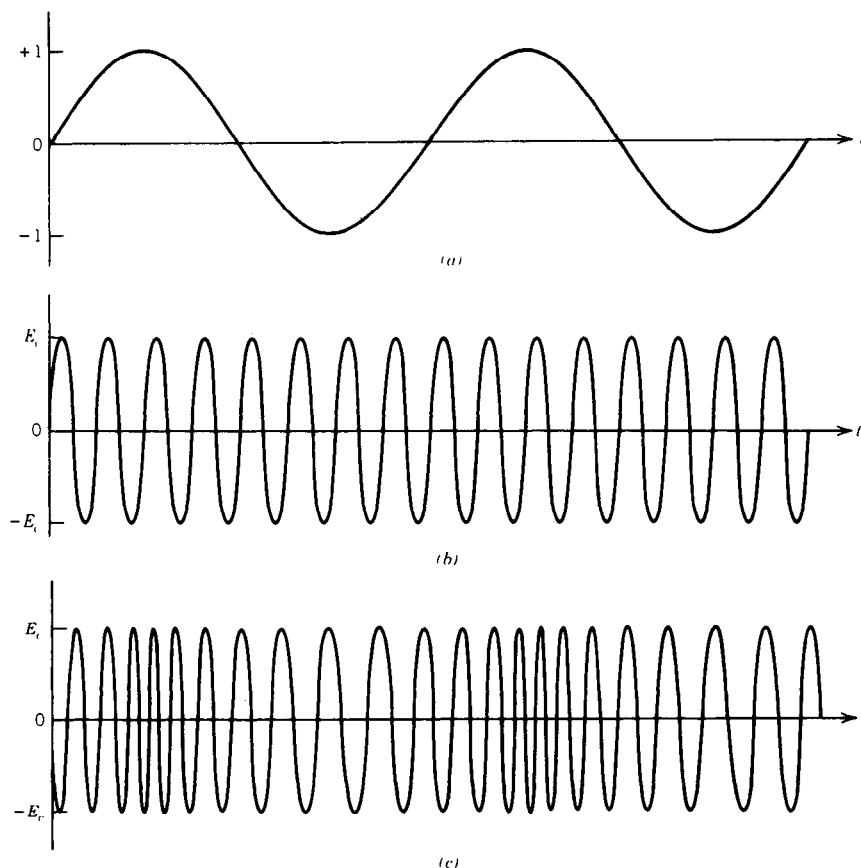


Fig. 48-13. Ilustración en el dominio del tiempo de la modulación angular mediante una señal de información sinusoidal única: a) señal de información, $g(t) = \sin \omega_m t$; b) señal portadora RF, $e_c(t) = E_c \sin \omega_c t$; c) señal portadora RF modulada angularmente, $e(t)_{PM} = E_c \sin [\omega_c t + \Delta\phi \sin \omega_m t]$; $e(t)_{FM} = E_c \sin \{\omega_c t + [(\Delta\omega/\omega_m) \cos \omega_m t]\}$.

$$e_{PM}(t) = E_c \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_K=-\infty}^{\infty} \prod_{i=1}^K [J_{n_i}(\Delta\theta_i)] \left\{ \begin{matrix} \text{sen} \\ \text{cos} \end{matrix} \right\} \left[\omega_c t + \sum_{i=1}^K n_i(\omega_i t + \phi_i) \right] \quad (48-23)$$

donde $\left\{ \begin{matrix} \text{sen} \\ \text{cos} \end{matrix} \right\}$ denota funciones seno o coseno y $\prod_{i=1}^K$ representa el producto de los términos $i = 1$ hasta K .

La adición de más tonos de modulación a una portadora modulada en fase toma potencia de los tonos de modulación existentes, así como de la portadora. Por consiguiente, antes de emplear técnicas de multicana- lización por división de frecuencia (subportadoras) para transferir información en un canal PM, es necesaria considerar los efectos de la distribución de potencia entre las diversas subportadoras cuando se elijan los índices de modulación y el número de subportadoras. En la figura 48-16 se ilustra la distribución de potencia del término de la portadora y los primeros términos de la banda lateral (J_1) de la densidad espectral de potencia para el caso de dos subportadoras PM.

Desmodulación de fase

Para la desmodulación de las señales PM suele utilizarse un proceso de recuperación de la portadora y detección del producto parecido al de la figura 48-17.

Dado que en la detección PM se utiliza el proceso de multiplicación, los resultados pueden considerarse sencillamente como una traslación del espectro modulado en fase a la banda base (fig. 48-18). Luego, la banda lateral ω_m (término J_1) puede filtrarse con un filtro pasabajas, y recuperarse la señal de la subportadora. La potencia en la señal recuperada es igual a la potencia total de la señal PM multiplicada por el cuadrado del término de Bessel de primer orden si no existe error de

fase entre la portadora de referencia en cuadratura y la portadora recibida; es decir (Painter y Hondros):

$$e_{o,PM}(t) = KE_c J_1(\Delta\theta) \text{sen } \omega_m t$$

donde $K = \cos \Delta\theta_c$. Entonces:

$$P_{SC} = \frac{E_c^2}{2} J_1^2(\Delta\theta)$$

que es la potencia contenida en el primer par de bandas laterales de la onda modulada en fase.

La relación de la potencia en el primer par de bandas laterales entre la potencia total de una onda modulada en fase para una señal subportadora sinusoidal se denomina pérdida o atenuación por modulación (ML, de *modulation loss*):

$$ML = \frac{2J_1^2(\Delta\theta)(E_c^2/2)}{E_c^2/2} = 2J_1^2(\Delta\theta)$$

El comportamiento del ciclo de recuperación de la portadora (rastreo de la portadora) está afectado por el número de tonos de modulación y sus respectivos índices de modulación. Este factor se denomina supresión de la portadora (CS, de *carrier suppression*) de la onda PM, y para la modulación de tono único es:

$$CS = J_0^2(\Delta\theta)$$

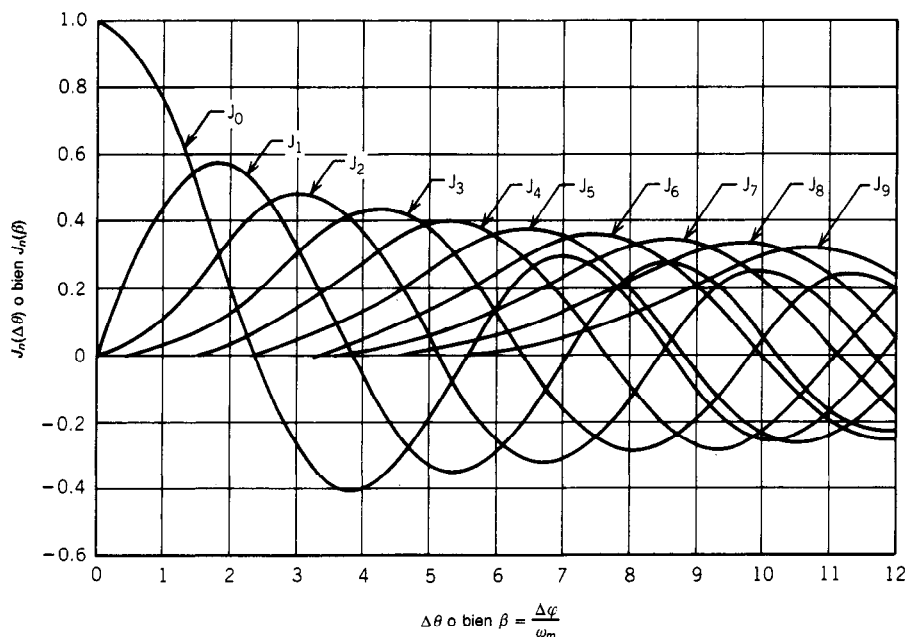


Fig. 48-14. Coeficientes de Bessel para las componentes espectrales de modulación de fase y de frecuencia (PM y FM).

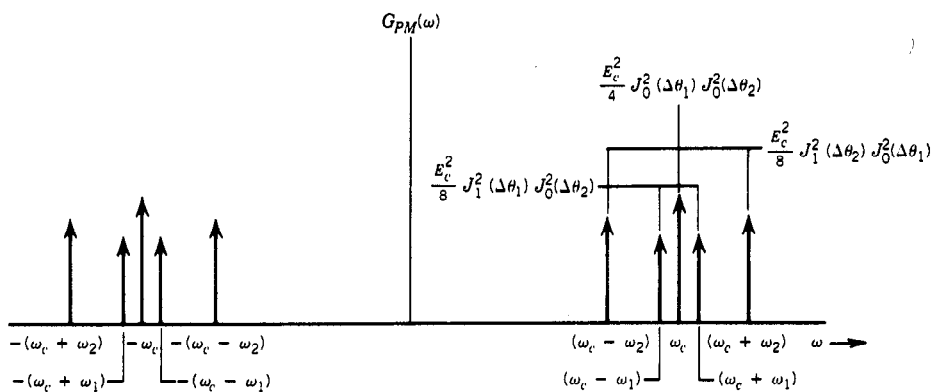
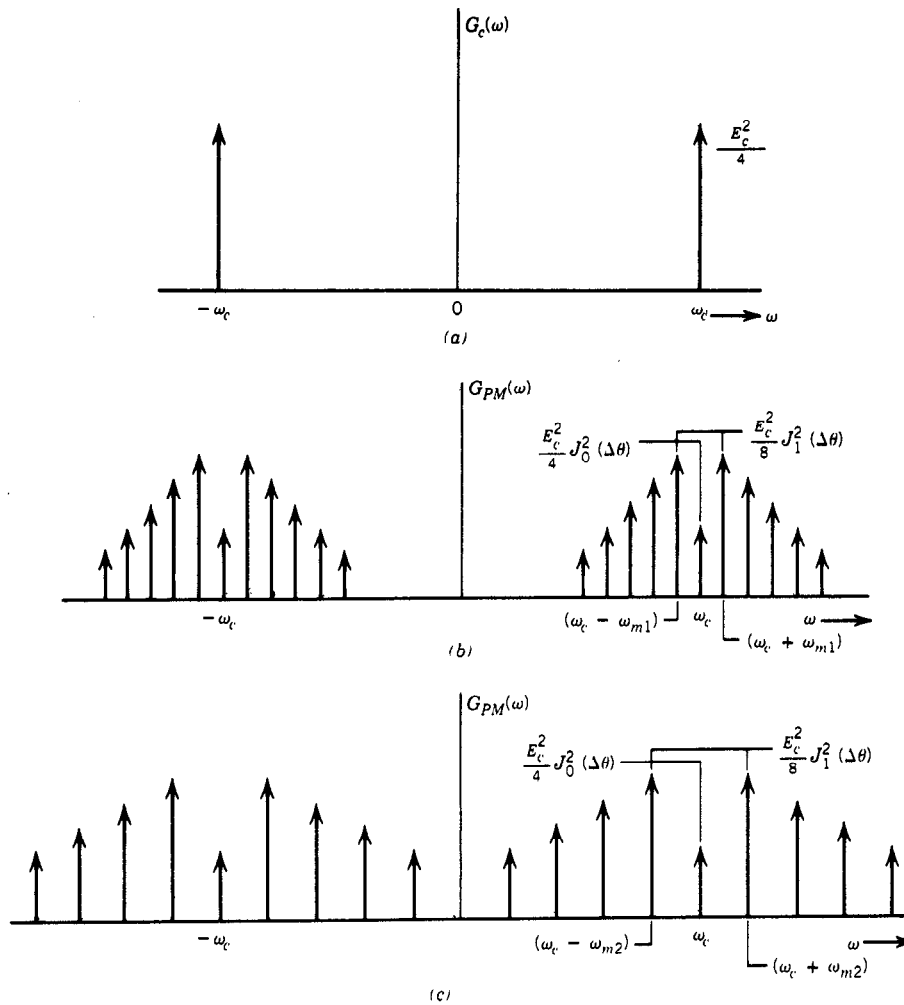


Fig. 48-16. Espectro de potencia parcial resultante de dos subportadoras a frecuencias ω_1 y ω_2 , con índices de modulación $\Delta\theta_1$ y $\Delta\theta_2$, respectivamente.

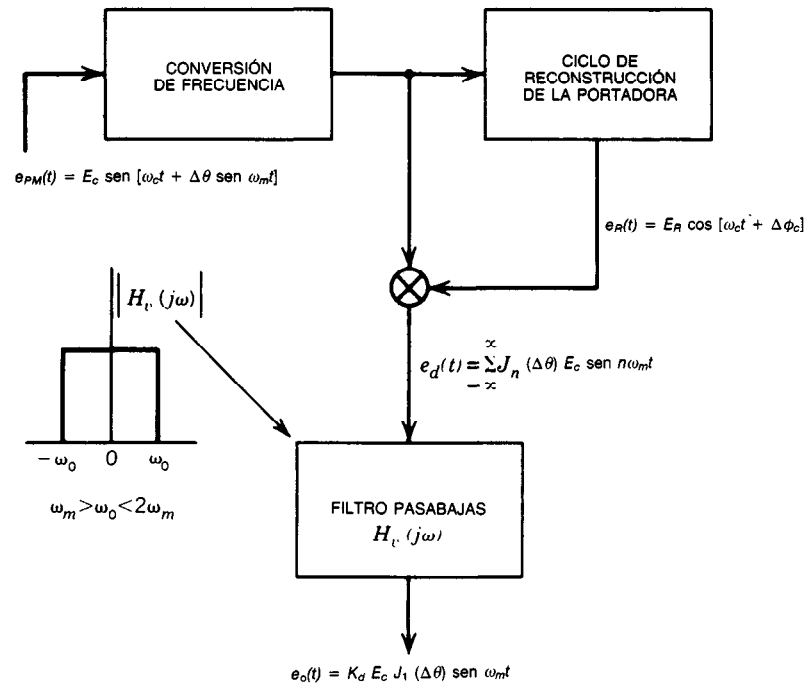


Fig. 48-17. Proceso de detección para una portadora modulada en fase por una sola senoide de frecuencia ω_m .

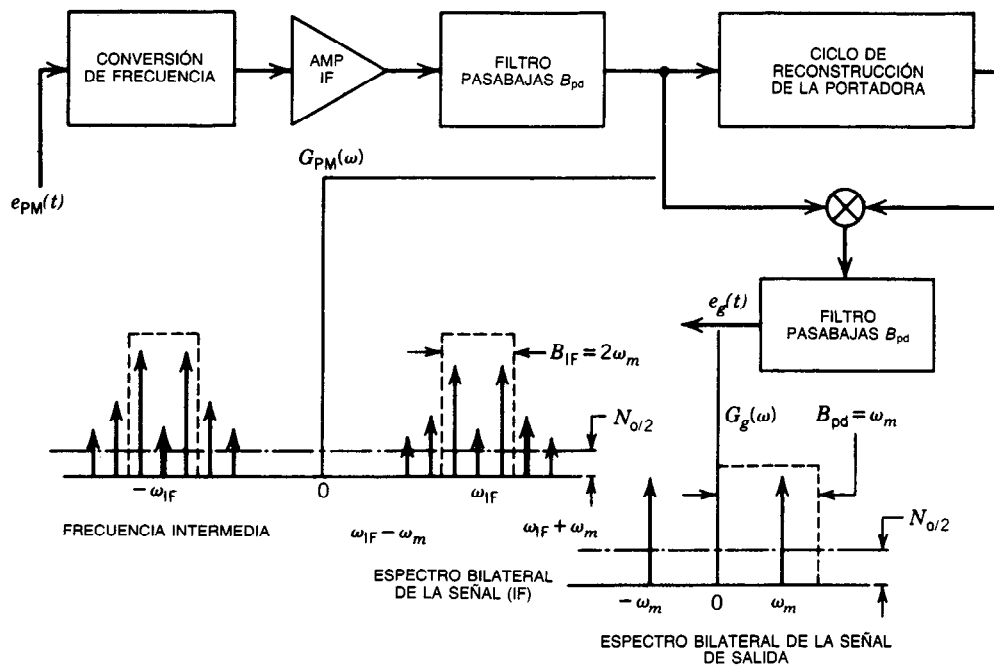


Fig. 48-18. Recepción y detección PM para modulación sinusoidal de tono único.

Tabla 48-3. Supresión de la portadora y pérdida por modulación contra $\Delta\theta$ para la modulación sinusoidal

$\Delta\theta$ (rad)	Supresión de la portadora (SP) $= J_0^2(\Delta\theta)$	Pérdida por modulación (PPM) $= 2J_1^2(\Delta\theta)$
0	1.0	0
0.2	0.980	0.018
0.4	0.922	0.076
0.6	0.832	0.162
0.8	0.716	0.272
1.0	0.585	0.388
1.2	0.450	0.496
1.4	0.320	0.588
1.6	0.207	0.650
1.8	0.115	0.676
2.0	0.050	0.666
2.2	0.017	0.618
2.4	6.3×10^{-6}	0.542
2.6	0.0094	0.442
2.8	0.034	0.336
3.0	0.068	0.230
3.2	0.102	0.136
3.4	0.133	0.064
3.6	0.153	0.018
3.8	0.162	0.0003
4.0	0.158	0.009

En la tabla 48-3 se enumeran la supresión de la portadora $J_0^2(\Delta\theta)$ y la pérdida por modulación $2J_1^2(\Delta\theta)$ para un canal único de subportadora sinusoidal como función del índice de modulación $\Delta\theta$.

Dado que solamente es necesario recuperar las primeras bandas laterales [es decir, los términos J_1

$(\Delta\theta) \sin \omega_m t$] de la señal modulada para la modulación de subportadora de tono único, el requisito de ancho de banda IF (frecuencia intermedia) del receptor es equivalente al de una onda modulada en amplitud de tono único.

Se supone que el ruido de entrada del receptor es blanco y gaussiano, con densidad espectral de ruido bilateral $N_0/2$. La relación entre la razón señal sobre ruido de salida (IF) y la de entrada (considerando solamente los términos fundamentales J_1 como potencia de la señal de entrada) es:

$$(S/N)_{o,PM} = \frac{E_c^2 J_1^2(\Delta\theta)}{2N_0\omega_m} \quad (48-24)$$

que es equivalente al 100% de rendimiento de AM, DSBSC y SSBSC basado en potencia de banda lateral constante. Sin embargo, debido al proceso de detección coherente implícito en la desmodulación PM, el rendimiento con bajas razones señal sobre ruido de entrada [es decir, $(S/N)_{ent} \ll 10$ dB] se mejora significativamente con respecto al de la detección AM no coherente ordinaria.

Modulación digital de fase

Una señal portadora modulada en fase por una señal digital (binaria) puede expresarse en el dominio del tiempo como:

$$e_{PM}(t) = E_c \sin [\omega_c t + \Delta\theta s(t)] \quad (48-25)$$

donde $s(t)$ se define como una función de conmutación digital del tiempo cuyos valores son ± 1 .

Si la transición de $+1$ a -1 y viceversa es instantánea (es decir, se supone ancho de banda infinito), entonces la expresión puede desarrollarse para obtener (Panneton):

$$e_{PM}(t) = E_c [\cos \Delta\theta \sin \omega_c t + s(t) \sin \Delta\theta \cos \omega_c t] \quad (48-26)$$

El espectro de potencia de una señal modulada en fase digitalmente puede observarse en la figura 48-19.

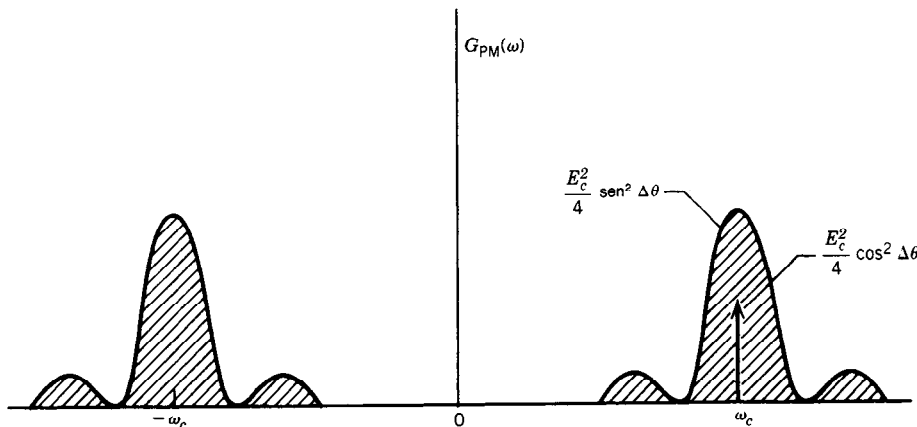


Fig. 48-19. Densidad espectral de potencia de una onda modulada en fase digitalmente [señal de información binaria sin regreso a cero (NRZ)].

Tabla 48-4. Supresión de la portadora y pérdida por modulación contra $\Delta\theta$ para la modulación de fase digital

$\Delta\theta$ (rad)	Supresión de la portadora (SP) $= \cos^2 \Delta\theta$	Pérdida por modulación (PPM) $= \sin^2 \Delta\theta$
0	1	0.0
0.2	0.961	0.0395
0.4	0.848	0.152
0.6	0.681	0.318
0.8	0.485	0.515
1.0	0.292	0.708
1.2	0.131	0.869
1.4	0.029	0.971
1.6	0.00085	0.999
1.8	0.052	0.948
2.0	0.173	0.827
2.2	0.346	0.654
2.4	0.544	0.456
2.6	0.734	0.266
2.8	0.888	0.112
3.0	0.980	0.019

La cantidad de potencia distribuida en las bandas laterales que representan la señal de información es:

$$P_{SB} = \frac{E_c^2}{2} \sin^2 \Delta\theta \quad (48-27)$$

Dado que la potencia total de la onda PM es $E_c^2/2$, la pérdida por modulación puede redefinirse para una señal de modulación digital como:

$$ML = \sin^2 \Delta\theta \quad (48-28)$$

donde $\Delta\theta$ es el índice de modulación PM ya definido.

La potencia restante de la portadora en la onda es:

$$P_c = \frac{E_c^2}{2} \cos^2 \Delta\theta \quad (48-29)$$

Así, la supresión de la portadora también puede definirse como:

$$CS = \cos^2 \Delta\theta \quad (48-30)$$

En la tabla 48-4 se presentan los valores de la pérdida por modulación (ML) y la supresión de la portadora (CS) como funciones de $\Delta\theta$. Cuando $\Delta\theta = \pi/2$ radianes, la portadora se suprime por completo:

$$CS = \cos^2 \frac{\pi}{2} = 0$$

y no ocurre pérdida por modulación:

$$ML = \sin^2 \frac{\pi}{2} = 1$$

Así, para un índice de modulación de $\pi/2$ radianes, suponiendo una función de modulación digital $s(t)$, la señal portadora se suprime por completo y toda la potencia aparece en las bandas laterales de la información. De este modo, la función de tiempo resultante se transforma en:

$$e_{PSK}(t) = E_c s(t) \cos(\omega_c t + \phi_0) \quad (48-31)$$

Este caso especial de PM digital (es decir, donde $\Delta\theta = \pm \pi/2$ radianes) se denomina manipulación por defasamiento o por desplazamiento de fase (PSK, de *phase shift keying*). Ya que las funciones temporales $e_{PSK}(t)$ y $e_{DSBSC}(t)$ son de la misma forma, es posible generar una señal PSK con un modulador lineal de producto configurado para modulación DSBSC. El espectro de la señal modulada resultante es sencillamente una traslación en frecuencia del espectro de información digital a la banda de frecuencia portadora en RF. En la figura 48-20 se ilustra el espectro de una señal PSK cuando la velocidad de transferencia de datos digitales es $\omega_R = 2\pi/T$.

Desmodulación digital de fase

Dado que la señal PM digital es equivalente a la modulación DSBSC, la detección puede efectuarse como se ilustra en la figura 48-21, utilizando una señal portadora reconstruida en cuadratura (defasada 90°) como la referencia del desmodulador (Lindsey). Como ya se observó, la salida del desmodulador puede entonces considerarse sencillamente el espectro de la señal IF trasladado a la banda base. La potencia de la señal contenida en las bandas laterales de la señal de información, suponiendo que no hay error de fase en el proceso de detección o pérdida de potencia debido a limitaciones en el ancho de banda IF, es $E_c^2 \sin^2 \Delta\theta/2$, donde $\Delta\theta$ es el índice de modulación PM.

Para $\Delta\theta = \pm \pi/2$ radianes (modulación PSK), la potencia de la señal de información es igual a la potencia total recibida. Suponiendo que el ruido del receptor es blanco y gaussiano con densidad espectral igual a $N_0/2$, la razón de salida de señal sobre ruido en el ancho de banda de la velocidad de transferencia de datos ω_R es:

$$(S/N)_{o,PSK} = \frac{E_c^2 \sin^2 \Delta\theta}{2N_0 \omega_R} \quad (48-32)$$

Para un sistema PSK práctico, el ancho de banda IF del receptor del canal PM que se requiere para pasar el 90% o más de la potencia de la señal debe variar de 1.5 a 6 veces la velocidad de transferencia de datos ω_R , dependiendo del formato del código digital. En las figuras 48-22 y 48-23 (Batson) se ilustra el porcentaje de la potencia de la señal pasado por un filtro IF con ancho de banda B_{IF} , suponiendo una señal portadora modulada por PSK. Los datos se proporcionan tanto para el formato de datos digitales sin regreso a cero (NRZ, de *nonreturn-to-zero*) como para el bifásico (bi- ϕ).

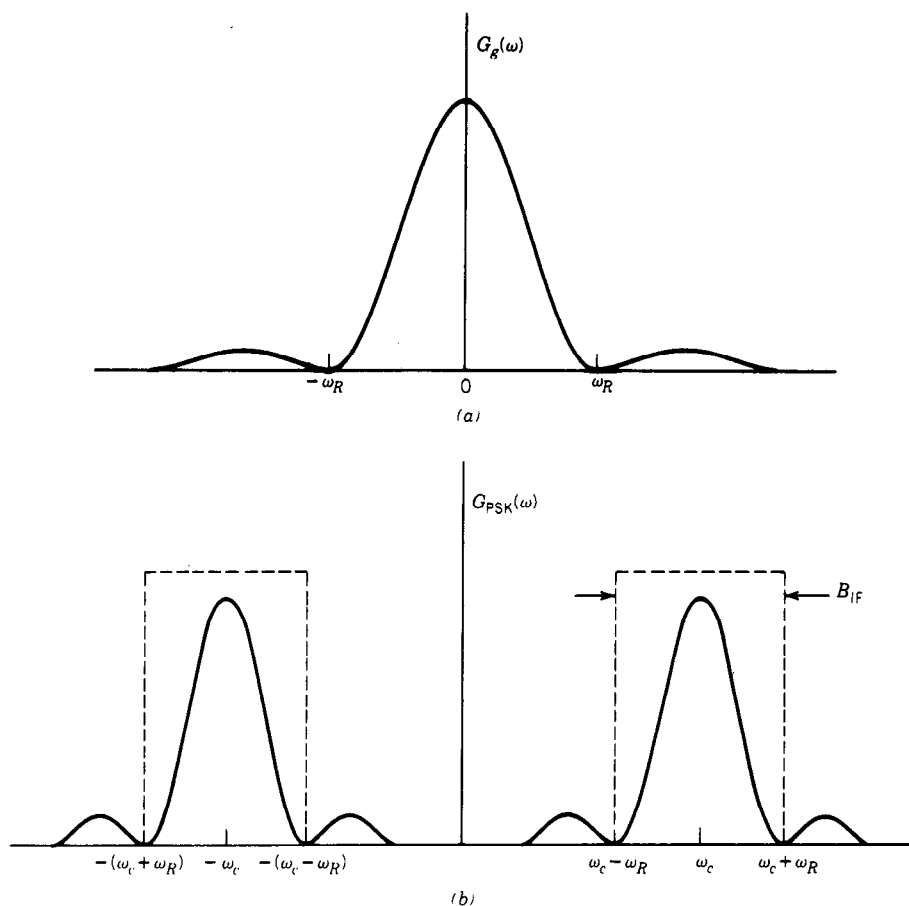


Fig. 48-20. Espectros de modulación digital mediante manipulación por desplazamiento de fase (PSK), suponiendo una señal de información sin regreso a cero (NRZ): *a*) espectro de potencia bilateral de la banda fundamental de la señal de información digital NRZ; *b*) espectro de potencia bilateral de la onda modulada por PSK.

Modulación de frecuencia

Para el caso de la modulación de frecuencia de onda continua, la expresión en el dominio del tiempo está dada por:

$$e_{FM}(t) = E_c \sin\left[\omega_c t + \Delta\omega \int_{-\infty}^t g(t) dt + \phi_0\right] \quad (48-33)$$

Como para el caso de la PM, el proceso de modulación en frecuencia es no lineal y da por resultado un espectro de ancho de banda infinito. Debido a la dificultad de establecer las características generales del espectro FM y los requisitos del ancho de banda de transmisión, solamente se analizarán unos cuantos tipos selectos de funciones de señal de modulación.

Modulación de frecuencia de onda sinusoidal. Para el caso especial en que la señal de modulación es una onda sinusoidal única:

$$g(t) = \Delta\omega \sin \omega_m t$$

La función de tiempo de la portadora modulada en frecuencia es:

$$e_{FM}(t) = E_c \sin\left(\omega_c t + \frac{\Delta\omega}{\omega_m} \cos \omega_m t + \phi_0\right) \quad (48-34)$$

El argumento de la onda modulada en frecuencia contiene la componente de fase dependiente del tiempo:

$$\theta(t) = \frac{\Delta\omega}{\omega_m} \cos \omega_m t \quad (48-35)$$

Así, la modulación en frecuencia por medio de un tono de onda sinusoidal único de una señal con frecuencia ω_m es equivalente a la modulación en fase de una portadora con una onda cosenoidal de frecuencia ω_m (radianes) e índice de modulación de fase $\Delta\theta = \Delta\omega/\omega_m$.

La onda modulada en frecuencia de la ecuación 48-34 puede desarrollarse en serie para obtener:

$$e_{FM}(t) = E_c \left[J_0 \left(\frac{\Delta \omega}{\omega_m} \right) \sin \omega_c t + \sum_{n=1}^{\infty} (-1)^{n+1} J_{2n-1} \left(\frac{\Delta \omega}{\omega_m} \right) \{ \cos[\omega_c + (2n-1)\omega_m]t + \cos[\omega_c - (2n-1)\omega_m]t \} \right. \\ \left. + \sum_{n=1}^{\infty} (-1)^n J_{2n} \left(\frac{\Delta \omega}{\omega_m} \right) \{ \sin[\omega_c + 2n\omega_m]t + \sin[\omega_c - 2n\omega_m]t \} \right] \quad (48-36)$$

El espectro de potencia de la onda FM con modulación sinusoidal es una serie infinita de pulsos separados simétricamente con respecto a la portadora por una distancia igual a la frecuencia de la señal de modulación ω_m y sus armónicos. El peso de cada impulso depende de la función de Bessel de orden adecuado para ese término específico de banda lateral y el índice de modulación de frecuencia $\beta = \Delta\omega/\omega_m$. Ya que el caso práctico no permite un ancho de banda de transmisión

$$BW_{FM} = 2(\Delta\omega + \omega_m) \quad (48-37)$$

que suele denominarse regla de Carson. Sin embargo, la aplicación de esta regla para la modulación en frecuencia con señales de información complejas no necesariamente da por resultado un rendimiento óptimo.

En la figura 48-24 se ilustra una onda FM modulada a medida que la frecuencia moduladora ω_m se hace variar y la desviación $\Delta\omega$ se mantiene constante.

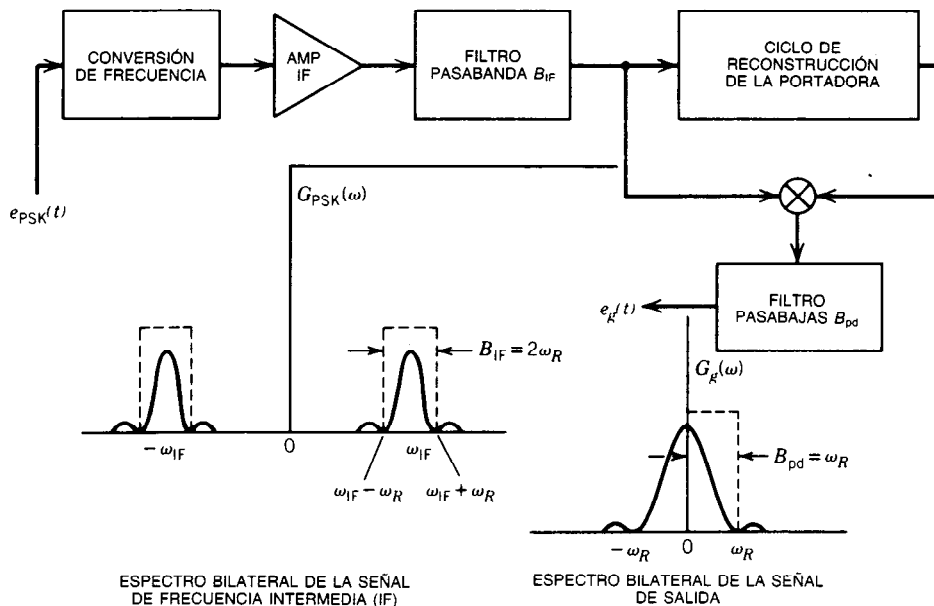


Fig. 48-21. Recepción y detección para la modulación en fase y la manipulación por desplazamiento de frecuencia (PSK) digitales.

infinito, ocurre algo de filtrado en el transmisor, en el receptor o en ambos. El ancho de banda de transmisión requerido por la portadora sinusoidal modulada en frecuencia depende del índice de modulación y de la frecuencia de la onda sinusoidal moduladora. Ha podido demostrarse (Painter y Hondros) que el ancho de banda que abarca todas las bandas laterales de amplitud significativa para este caso especial es equivalente a:

En el caso de la modulación de ondas sinusoidales múltiples (no relacionadas armónicamente), el espectro es de ancho de banda infinito, así como para la PM multiton. Las componentes espectrales se presentan a frecuencias correspondientes a la suma y la diferencia de cada tono de modulación y sus armónicos, con todos los demás tonos de modulación y sus armónicos. La expresión en el dominio del tiempo para la onda modulada multiton es (Giacoletto):

$$e_{FM}(t) = E_c \sum_{n_1=-\infty}^{\infty} \dots \sum_{n_K=-\infty}^{\infty} \prod_{i=1}^K \left[J_{n_i}(\beta_i) \left\{ \begin{matrix} \sin \\ \cos \end{matrix} \right\} \right] \left[\omega_c t + \sum_{i=1}^K n_i(\omega_i t + \phi_i) \right] \quad (48-38)$$

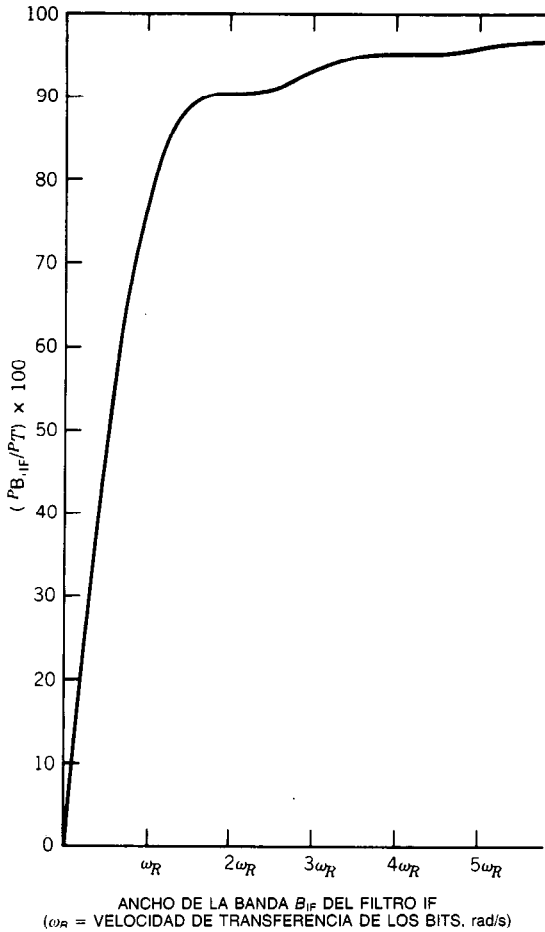


Fig. 48-22. Porcentaje del total de la potencia de la señal pasado por un filtro de frecuencia intermedia (IF) para una señal modulada por un desplazamiento de fase (PSK) [se supone una señal de información sin regreso a cero (NRZ)].

Los requisitos de ancho de banda para la FM multiton no coherente pueden deducirse evaluando las funciones de Bessel. Ya que la potencia de onda FM es constante e igual a la potencia portadora, es obvio que la potencia en cada componente de banda lateral decrece a medida que aumenta el número de tonos de modulación. Más allá de cierto intervalo de frecuencias con respecto a la portadora, la potencia de la banda lateral se vuelve insignificante.

Si β_i es grande para cada tono, el ancho de banda requerido para pasar el 90% o más de la potencia de banda lateral en el receptor IF es aproximadamente (Panter):

$$BW_{FM} = 2 \sum_{i=1}^K \Delta\omega_i \quad (48-39)$$

Si la relación de desviación es pequeña, $\beta_i \ll 1$, el espectro consistirá en un par de bandas laterales a la frecuencia fundamental ω_i de cada tono de modulación.

Por consiguiente, el ancho de banda requerido para pasar por lo menos el 90% de la potencia de la señal es (Panter):

$$BW_{FM} = 2\omega_J \quad (48-40)$$

donde ω_J es la mayor frecuencia del tono de modulación.

Para valores intermedios de β , el ancho de banda requerido está dado por (Panter):

$$BW_{FM} = 2(\Delta\omega_p + 2\omega_J) \quad (48-41)$$

donde $\Delta\omega_p$ es la desviación compuesta de la frecuencia pico del sistema y ω_J es la mayor frecuencia del tono de modulación.

Modulación digital de frecuencia. Dos casos de FM digital de interés son la FM de fase continua, que es parecida al proceso de modulación en frecuencia analizado hasta el momento, y la manipulación por desplazamiento de frecuencia (FSK, de *frequency shift keying*), que es equivalente a la conmutación entre los osciladores de las frecuencias ω_1 y ω_2 , siempre en función de la señal de información digital que se ha de transmitir.

La modulación de fase continua con una onda cuadrada o un proceso digital binario puede expresarse como:

$$e_{FM}(t) = E_c \sin[\omega_c t + \Delta\omega s(t)] \quad (48-42)$$

donde $s(t)$ representa la función que asume solamente los valores +1 o -1, dependiendo del símbolo de información que se esté transmitiendo.

La determinación del espectro requiere el conocimiento de los coeficientes del desarrollo en serie de Fourier de $e_{FM}(t)$.

Si $s(t)$ es una onda cuadrada o una sucesión repetitiva de 1 y 0, entonces la onda de forma FM puede expresarse como (Panter):

$$e_{FM}(t) = \sum_{n=-\infty}^{\infty} \bar{F}_n \cos(\omega_c t + n\omega_0 t) \quad (48-43)$$

donde ω_0 es la frecuencia fundamental de la onda cuadrada de modulación y:

$$\bar{F}_n = \frac{1}{2} \left\{ \frac{\sin[(\beta - n)\pi/2]}{(\beta - n)\pi/2} + (-1)^n \frac{\sin[(\beta + n)\pi/2]}{(\beta + n)\pi/2} \right\} \quad (48-44)$$

donde $\beta = \Delta\omega/\omega_0$.

En la figura 48-25 se ilustra el espectro modulado; ahí la razón de desviación se incrementa manteniendo constante $\Delta\omega$ y reduciendo ω_0 . Como puede observarse, la energía espectral tiende a concentrarse en la frecuencia equivalente a la desviación de la frecuencia $\Delta\omega$. La distribución espectral FM tiende a seguir la distribución en el tiempo de la señal moduladora. Cuanto más tiempo permanezca la señal en una localización de

frecuencia portadora específica, más energía espectral se presenta en esa banda del espectro de la señal modulada. A medida que β tiende a infinito con $\Delta\omega$ fijo, la velocidad de transmisión de información tiende a cero en el límite. Así, cuando $\beta = \infty$, la señal de modulación está en $+\Delta\omega$ y en $-\Delta\omega$ durante un periodo infinitamente grande, lo que da por resultado que una mitad de la potencia esté localizada en $\omega_c + \Delta\omega$ y la otra mitad en $\omega_c - \Delta\omega$.

Señalización FSK. En el caso especial de la manipulación por desplazamiento de frecuencia, la señal modulada se forma mediante la conmutación entre dos osciladores a las frecuencias ω_1 y ω_2 . Dado que los osciladores no son coherentes, la fase de la onda modulada resultante no es continua; es decir, ocurre una discontinuidad en el tiempo de conmutación. En la figura 48-26 se ilustra la técnica de la modulación FSK. El espectro de la señal resultante consta de dos funciones de pulso a las frecuencias de símbolos, moduladas por el proceso aleatorio de información binaria que se ilustra en la figura 48-27.

En virtud de que la señal FSK no es continua en fase, puede recuperarse por medio del filtrado individual de cada una de las frecuencias de símbolos en el receptor. Por tanto, el ancho de banda real requerido para la recepción de FSK es menor que $\omega_1 + \omega_2$, como puede inferirse a partir de la figura 48-27.

Desmodulación de frecuencia. En la figura 48-28 se ilustra la recepción y detección de FM. Si el ruido IF en la salida del filtro pasabanda B_{IF} es blanco y gaussiano

con densidad espectral bilateral del ruido $N_0/2$ y la relación de señal (portadora) sobre ruido es grande (es decir, $C/N > 10$ dB), el espectro de potencia del ruido de salida (fig. 48-29) está dado por:

$$G_{n0}(\omega) = \begin{cases} \frac{K_d^2 \omega^2}{4\pi^2 E_c^2} N_0 & |\omega| < 2\pi \left[\frac{B_{IF}}{2} \right] \\ 0 & |\omega| > 2\pi \left[\frac{B_{IF}}{2} \right] \end{cases} \quad (48-45)$$

Obsérvese que el espectro del ruido de salida depende de ω^2 , lo que produce una forma parabólica. Entonces, el ruido de salida de la banda de base se obtiene integrando la función de transferencia del filtro pasabajos de posdetección. Suponiendo que la sensibilidad del discriminador está dada por K_d y que el filtro de salida es un filtro pasabajos ideal con respuesta:

$$|H(j\omega)|^2 = \begin{cases} 1 & |f| < B_{pd} \\ 0 & |f| > B_{pd} \end{cases} \quad (48-46)$$

entonces, suponiendo que la portadora de entrada no está modulada por una señal de información [es decir, que $g(t) = 0$], la potencia del ruido de salida es:

$$P_n = \left(\frac{K_d}{E_c} \right)^2 \frac{2N_0 B_{pd}^3}{3} \quad (48-47)$$

donde E_c = amplitud de la portadora a la entrada del discriminador

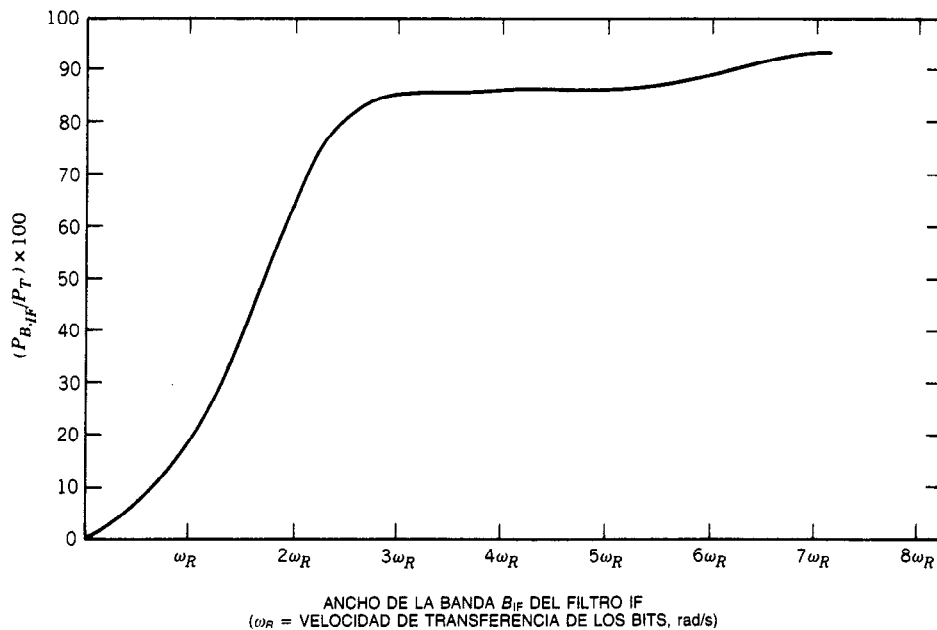


Fig. 48-23. Porcentaje del total de la potencia de la señal pasado por un filtro de frecuencia intermedia (IF) para una onda modulada por desplazamiento de fase (PSK) (suponiendo una señal de información de dos fases).

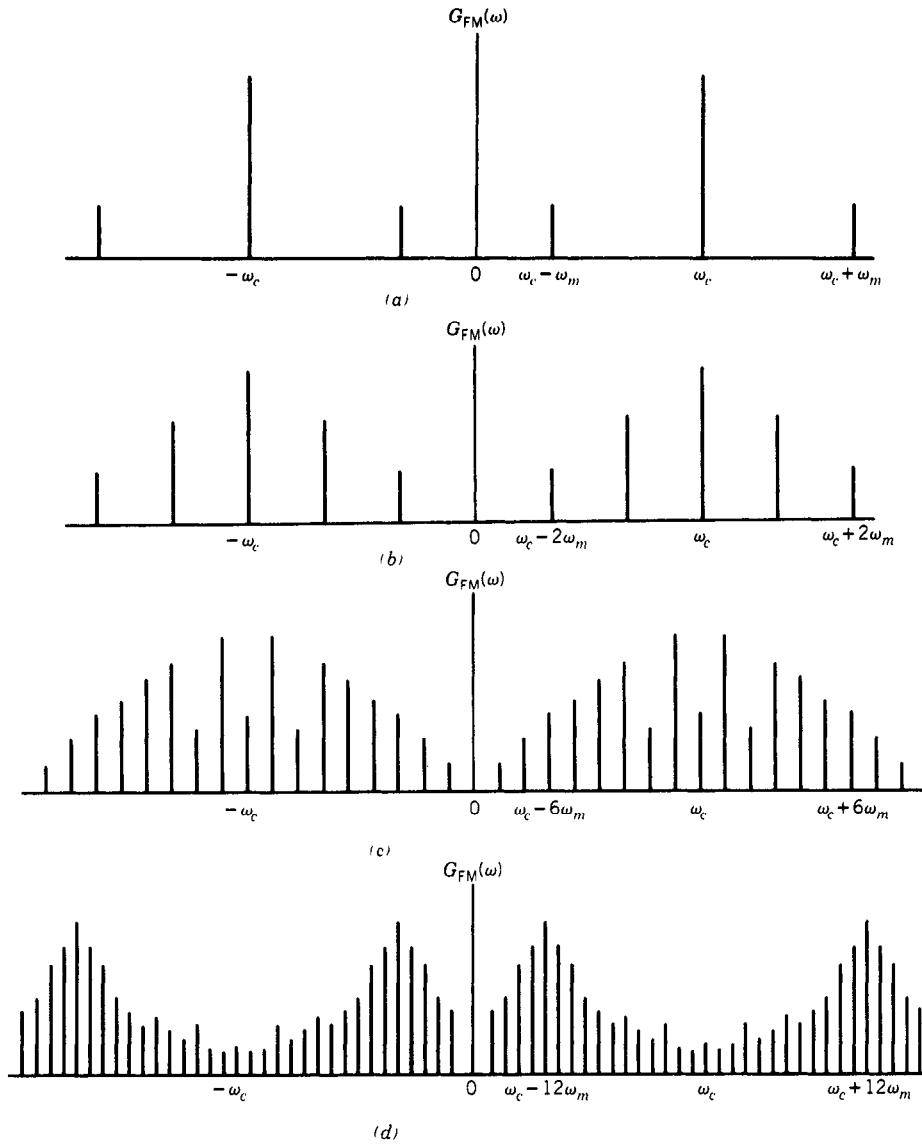


Fig. 48-24. Espectros de la potencia de modulación FM para una señal de modulación de tono único ($\Delta\omega$ constante): a) $\Delta f = 24$ kHz, $f_m = 48$ kHz, $\beta = 0.5$; b) $\Delta f = 24$ kHz, $f_m = 24$ kHz, $\beta = 1.0$; c) $\Delta f = 24$ kHz, $f_m = 4$ kHz, $\beta = 6.0$; d) $\Delta f = 24$ kHz, $f_m = 2$ kHz, $\beta = 12$.

$N_0/2$ = densidad espectral bilateral del ruido

B_{pd} = ancho de banda unilateral del filtro ideal de posdetección

Si la señal de información es $g(t)$, entonces la potencia de la señal de salida del discriminador es:

$$P_s = \left(\frac{K_d}{2\pi} \right)^2 \Delta\omega^2 [\overline{g^2(t)}] \quad (48-48)$$

y la relación de señal sobre ruido fuera del discriminador para relaciones IF de portadora sobre ruido mayores que 10 dB puede expresarse como:

$$(S/N)_{o,FM} = \frac{3E_c^2 \Delta f^2 \overline{g^2(t)}}{2N_0 (B_{pd})^3} \quad (48-49)$$

Por otra parte, la relación de entrada de portadora sobre ruido en el ancho de banda IF está determinada por:

$$\text{CNR}_i = \frac{E_c^2}{2N_0 B_{IF}} \quad (48-50)$$

Si la señal de información es una onda sinusoidal, entonces $\overline{g^2(t)} = 1/2$ y:

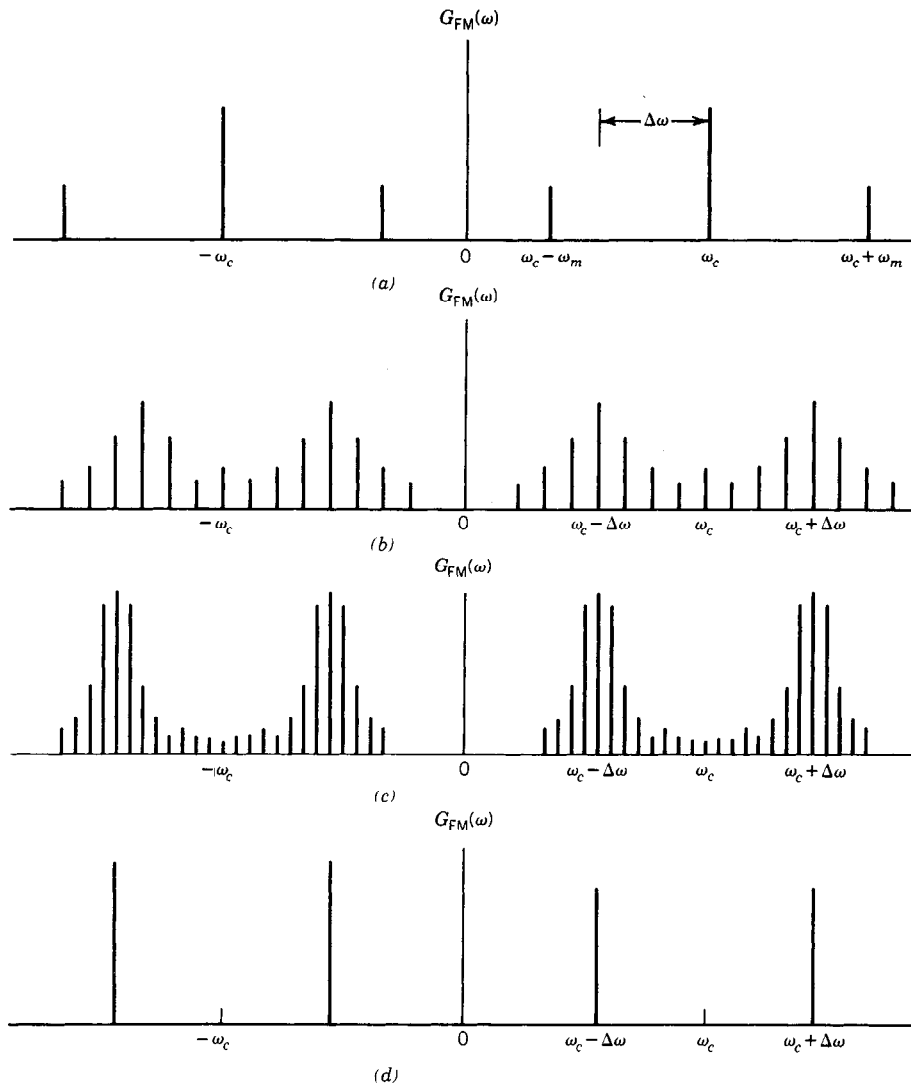


Fig. 48-25. Espectros de la potencia de modulación FM para señales de modulación de onda cuadrada ($\Delta\omega$ constante): a) $\beta < 1$; b) $5 < \beta < 10$; c) $10 < \beta < 25$; d) $\beta = \infty$.

$$(S/N)_{o,FM} = \frac{3}{2} \frac{\Delta f^2 B_{IF}}{B_{pd}^3} \text{CNR}_i \quad (48-51)$$

La mejora en la razón de salida de señal sobre ruido respecto a la razón de entrada de portadora sobre ruido se define como:

$$I_{FM} = \frac{3}{2} \Delta f^2 \frac{B_{IF}}{B_{pd}^3} \quad (48-52)$$

Si el filtro pasabajas de salida es ideal y el ancho de banda es equivalente a la frecuencia de la señal de modulación, ω_m , y si el ancho de banda IF se hace igual a dos veces la frecuencia de la señal de modulación, $2\omega_m$, entonces es posible comparar de manera directa con el rendimiento de señal sobre ruido de la AM.

Para estas condiciones, y suponiendo que la operación está por encima del umbral del discriminador, el factor de mejora es:

$$I_{FM} = 3 \left(\frac{\Delta f}{f_m} \right)^2 \quad (48-53)$$

y dado que $\Delta f/f_m$ se definió previamente como el índice de modulación β , la mejora de la FM sobre la AM a menudo se define como $3\beta^2$. Esto es cierto sólo para las muy especiales condiciones de posdetección y ancho de banda IF equivalentes a los requisitos del canal AM.

Cuando la señal de modulación se compone de una o más subportadoras sinusoidales, el filtrado de posdetección que suele utilizarse es el de la selección de la

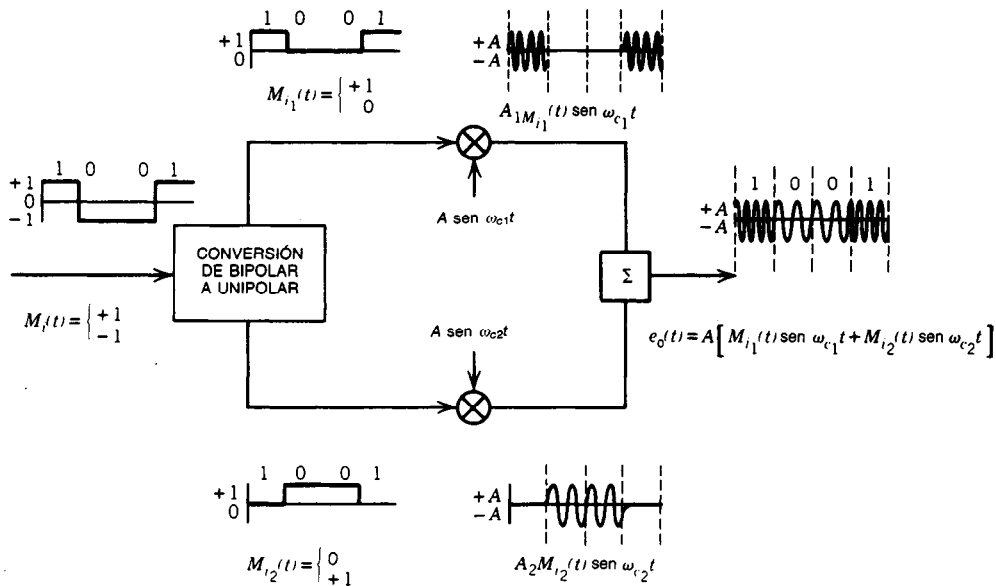


Fig. 48-26. Representación funcional de la manipulación por desplazamiento de frecuencia ortogonal FSKO (coherente) [suponiendo que $M_i(t)$ es una señal digital sin regreso a cero (NRZ)].

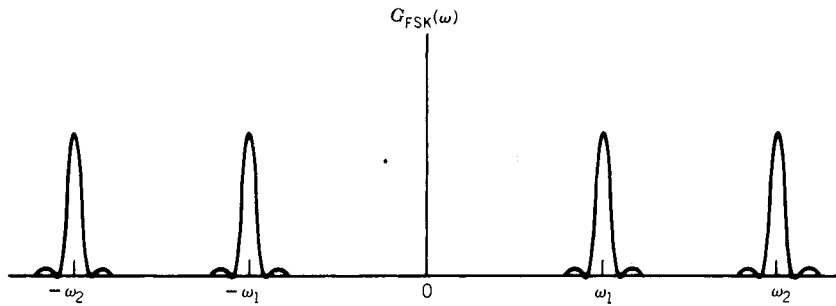


Fig. 48-27. Densidad espectral de potencia, señal de manipulación por desplazamiento de frecuencia (FSK).

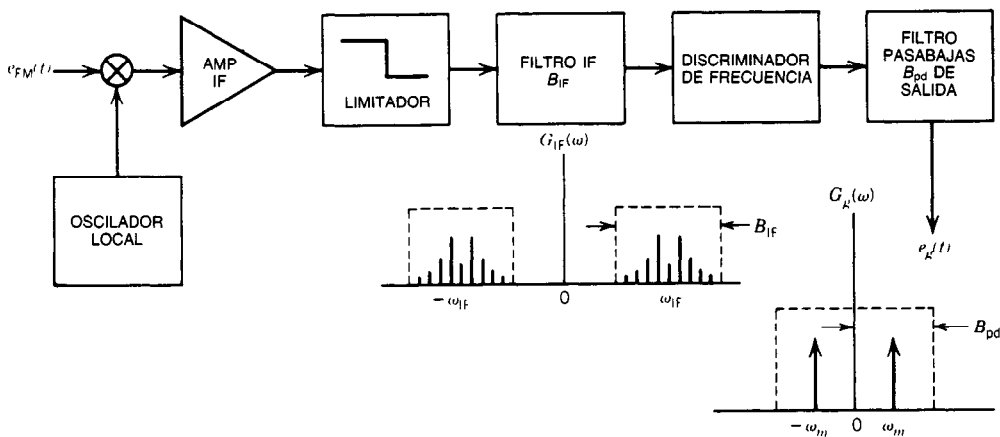


Fig. 48-28. Sistema típico de recepción y detección por FM (señal de modulación de tono único).

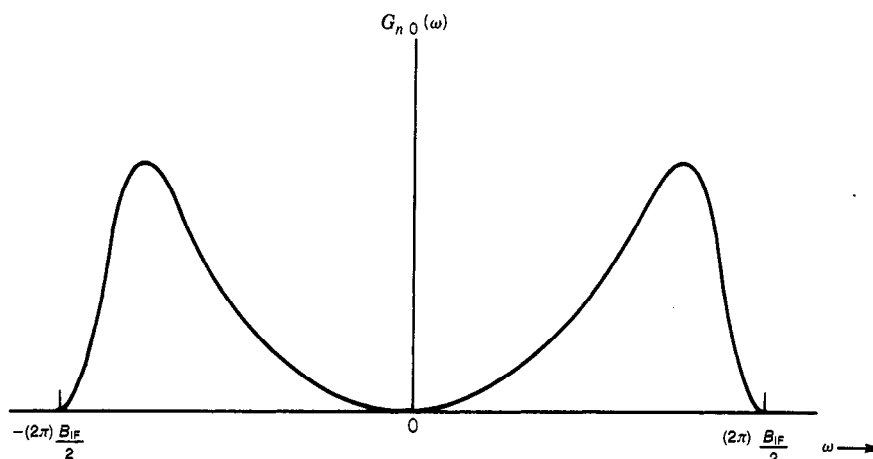


Fig. 48-29. Densidad espectral de la potencia del ruido de salida del discriminador de FM.

banda de paso de las señales subportadoras individuales, como se ilustra en la figura 48-30. En este caso el factor de mejora de señal sobre ruido, cuando se opera por encima del umbral del discriminador, es:

$$I_{FM,BPF} = \frac{3}{2} \frac{\Delta f^2 B_{IF}}{(f_U^3 - f_L^3)} \quad (48-54)$$

donde f_U es la frecuencia de corte superior de potencia mitad del filtro pasabanda (BPF, de *bandpass filter*) de posdetección, y f_L es la frecuencia de corte inferior de dicho filtro.

Las relaciones de mejora de FM definidas hasta el momento se han obtenido empleando aproximaciones válidas solamente en condiciones de elevada razón IF de portadora sobre ruido ($CNR_i > 10$ dB). Por debajo de cierto valor crítico de la CNR (de *carrier-to-noise ratio*), denominado nivel crítico, el funcionamiento del sistema FM se deteriora con rapidez. El análisis del

funcionamiento por debajo del nivel crítico es algo complicado. La expresión resultante para la razón de señal sobre ruido es (Shilling y Taub):

$$(S/N)_{o,FM} = \frac{3\rho^2 \Delta f^2 (B_{IF}/B_{pd}^3) CNR_i}{\sqrt{3} (B_{IF}/B_{pd})^2 CNR_i (1 - \text{erf} \sqrt{CNR_i}) + 1} \quad (48-55)$$

En esta expresión, ρ es la relación de valor cresta sobre valor rms (eficaz) de la señal moduladora, y $\text{erf} \sqrt{CNR_i}$ es la función gaussiana de error. Obsérvese que para una señal moduladora sinusoidal $\rho = 1/\sqrt{2}$ y que para una CNR_i fuerte la función de error tiende a la unidad. Entonces, la ecuación pasa a ser:

$$(S/N)_{o,FM} = \frac{3}{2} \Delta f^2 \frac{B_{IF}}{B_{pd}^3} CNR_i \quad (48-56)$$

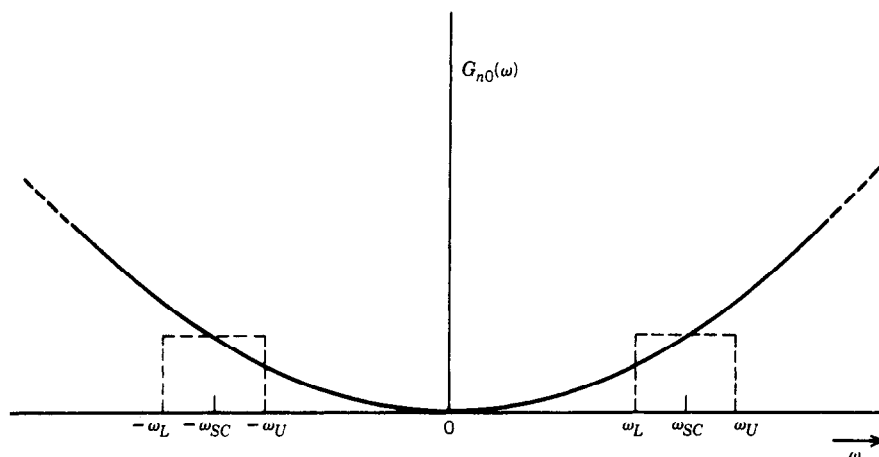


Fig. 48-30. Filtrado de posdetección pasabanda del discriminador de FM.

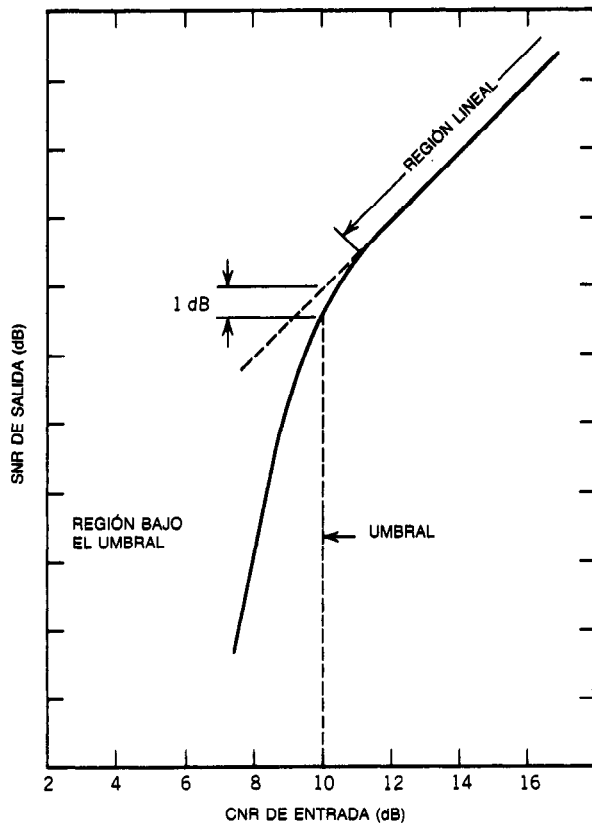


Fig. 48-31. Comportamiento típico de la razón de salida de señal sobre ruido (SNR) en función de la razón de entrada de portadora sobre ruido (CNR) para un discriminador de FM, en el que se ilustra el fenómeno de umbral.

como ya se analizó. En la figura 48-31 se presentan las características de función de transferencia de la razón de señal sobre ruido de la banda base para un discriminador típico FM; se muestra el efecto de umbral. El nivel umbral de la CNR, se define como la CNR, para la cual la razón de salida de señal sobre ruido $(S/N)_{o,FM}$ se ha desviado 1 dB respecto a la proyección en línea recta de la curva de mejora de la FM de señal fuerte. El ruido de salida de la banda base se vuelve impulsivo cuando se presenta la condición de nivel umbral, produciendo picos en la señal del ruido de salida. Estos picos o pulsos suelen denominarse chasquidos, y explican la rápida degradación en la razón de salida de señal sobre ruido que ocurre por debajo del nivel umbral. Este efecto de umbral explica por qué no es posible incrementar de manera indefinida la razón de salida de señal sobre ruido de los sistemas FM simplemente incrementando la desviación de la frecuencia Δf , ya que también ocurre un incremento correspondiente en el ancho de banda IF y por tanto una reducción en la razón de entrada de portadora sobre ruido.

Desmodulación FSK. En el proceso de desmodulación en frecuencia de una señal FSK pueden utilizarse téc-

nicas de detección coherente o no coherente, como se ilustra en la figura 48-32. Se ha demostrado que el rendimiento teórico óptimo en términos de probabilidad de error en los bits para este tipo de señalización es 3 dB inferior al de un canal PSK que opere a la misma velocidad de transmisión de datos (Batson).

La probabilidad de error para un sistema FSK coherente está dada por (Batson):

$$P_e = \frac{1}{2} \operatorname{erfc} \sqrt{\frac{(1 - \rho)E_b}{2N_0}} \quad (48-57)$$

donde E_b = energía por tiempo de bit

N_0 = densidad espectral unilateral de ruido

erfc = función error complementaria

ρ = coeficiente de correlación normalizado

El coeficiente de correlación normalizado se define:

$$\rho = \int_0^T S_1(t)S_2(t)dt \quad (48-58)$$

donde $S_1(t)$ y $S_2(t)$ son los símbolos transmitidos de energía unitaria. Puede demostrarse que para el caso

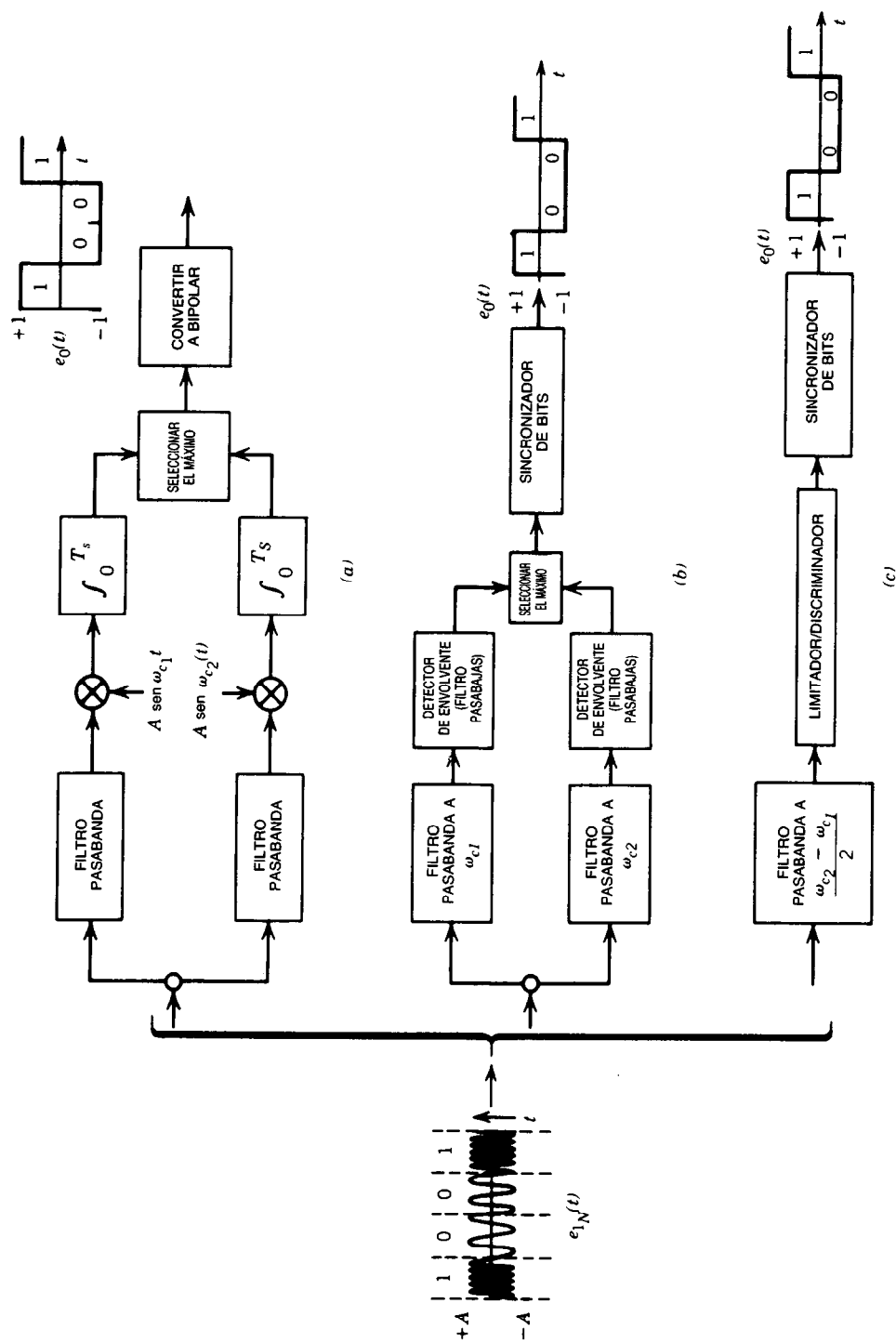
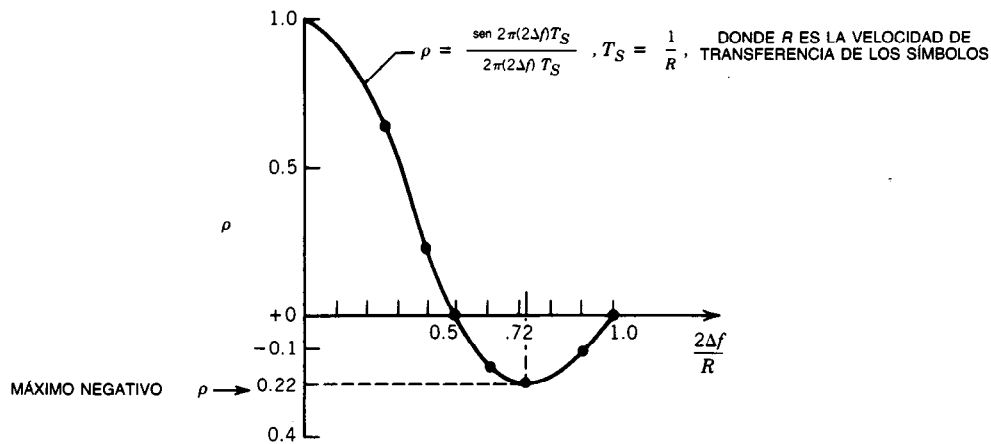


Fig. 48-32. Ilustración funcional de los procesos de desmodulación coherente e incoherente: a) detección coherente; b) detección de la envolvente (incoherente); c) detección por el discriminador (incoherente).



$$P_{E_o} = \frac{1}{2} \operatorname{erfc} \sqrt{\frac{.5 E_b}{N_o}} \rightarrow \text{FUNCIONAMIENTO ORTOGONAL} \left[\rho = 0, \frac{2\Delta f}{R} = .5 \right]$$

$$P_{E_n} = \frac{1}{2} \operatorname{erfc} \sqrt{\frac{.61 E_b}{N_o}} \rightarrow \text{MEJOR FUNCIONAMIENTO} \left[\rho = -.22, \frac{2\Delta f}{R} = .72 \right]$$

P_{E_n} REPRESENTA LA COTA DE FUNCIONAMIENTO EN FSK COHERENTE 2.2 dB PEOR QUE LA PSK COHERENTE

Fig. 48-33. Coeficiente de correlación y probabilidades resultantes de errores en los bits para modulación FM digital de fase continua [con señal de modulación digital sin regreso a cero (NRZ)].

de la transmisión FM de fase continua el coeficiente de correlación óptimo es diferente de cero (Seyl y colaboradores). En la figura 48-33 se muestra gráficamente el coeficiente de correlación suponiendo una señal de modulación digital sin regreso a cero (NRZ).

Como puede observarse en la figura, la desviación de la frecuencia de cresta para el funcionamiento ortogonal (es decir, $\rho = 0$) es $\Delta f = 0.25R$, donde R es la velocidad de transmisión de datos de la modulación NRZ. Sin embargo, el mejor rendimiento (óptimo), $\rho = -0.22$ y la desviación de la frecuencia de cresta requerida es $\Delta f = 0.35R$. Por consiguiente, sólo es posible obtener un rendimiento óptimo 2.2 dB peor que la PSK óptima.

Otro parámetro que afecta el rendimiento óptimo es el ancho de banda del filtro IF del receptor. Se han utilizado resultados de simulaciones y numéricos a fin de establecer la razón óptima de ancho de banda del filtro IF, BW_{IF} , sobre velocidad de transmisión de datos para la señalización digital FM de fase continua cuando se utiliza el Δf óptimo. Las razones óptimas $\Delta f/R$ y BW_{IF}/R para los formatos de datos NRZ y bifásicos están dadas en la tabla 48-5.

48.1.3 Modulación de pulsos

En los sistemas de modulación de pulsos, los datos se representan por medio de una serie de pulsos periódicamente recurrentes. La señal de información $g(t)$

es transportada haciendo variar algún parámetro asociado con el tren de pulsos. Dado que se dispone de varios parámetros que es posible variar, es factible un gran número de métodos de modulación. El método de modulación de pulsos más sencillo es la modulación de pulsos en amplitud o modulación de amplitud de pulsos (PAM, de *pulse amplitude modulation*), en el que se utilizan muestras de la señal de información para modular las amplitudes de los pulsos consecutivos de la portadora. En vez de utilizarse para modular la amplitud de pulsos, las muestras podrían emplearse para modular el ancho o duración de los pulsos y así obtener modulación de pulsos en ancho (PWM, de *pulse width modulation*) o modulación de pulsos en duración (PDM, de *pulse duration modulation*). En este caso la información se transporta en los cruces por

Tabla 48-5. Parámetros óptimos para la FM de fase continua con señales de información digital

Formato de los datos	$\Delta f/R^a$	B^b/R
Sin regreso a cero (NRZ)	0.36	1.4
Bifásico	0.62	2.7

Fuente. Tomado de Seyl, Batson y Smith, con permiso.

^a R , velocidad de transferencia de los datos de la señal de información.

^b B , ancho de banda de la frecuencia intermedia (FI) del receptor.

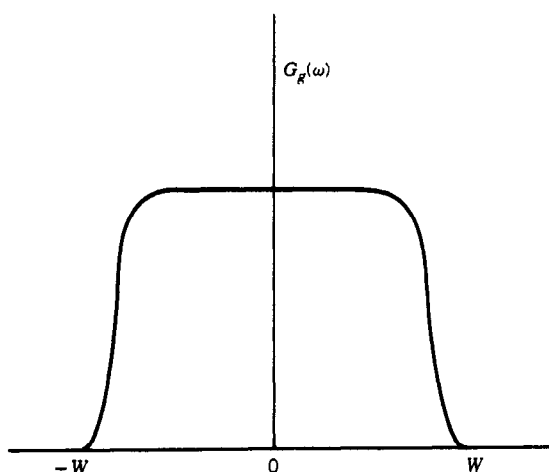


Fig. 48-34. Espectro idealmente limitado en banda de $g(t)$.

cero de los flancos de ascenso y descenso de los pulsos. Otro empleo sería la modulación de pulsos en posición (PPM, de *pulse position modulation*), en la cual se hace variar la duración del pulso con respecto al tiempo de repetición (referencia) de los pulsos regularmente espaciados. Por último, es posible hacer variar la frecuencia del tren de pulsos a fin de obtener modulación de pulsos en frecuencia (PFM, de *pulse frequency modulation*).

En un método más complicado para la modulación de pulsos se incorpora una técnica que consiste en codificar muestras individuales de la amplitud de la señal de información y transmitir la información codificada a fin de representar cada tamaño de muestra. El proceso de muestrear y codificar se denomina modulación de pulsos por código (PCM, de *pulse code modulation*). Las muestras de la señal de información en la PCM deben convertirse en valores numéricos discretos preasignados; por tanto, cada muestra debe almacenarse en el depósito de código que represente su valor real con mayor precisión. El proceso de digitalizar una señal de información se denomina cuantización, y da por resultado algo de ruido extra (ruido de cuantización) en los datos recuperados. Otra forma de modulación de pulsos es la modulación en delta, en la cual unos símbolos discretos relacionados con aproximaciones previas a la muestra se transmiten conforme a un algoritmo específico. Un algoritmo parecido se aplica en el receptor a fin de recuperar la señal de información original. La modulación en delta se utiliza a menudo para la digitalización del habla, donde las velocidades de transmisión de datos son limitadas. Dado que en todos los esquemas de modulación de pulsos es necesario muestrear la señal de información a intervalos de tiempo regulares, los conceptos de muestreo de señales y de velocidad requerida de muestreo son importantes.

Teorema del muestreo

El teorema del muestreo establece el número de muestras en el tiempo de la función de tiempo de la infor-

mación $g(t)$ requerido para la reproducción completa de $g(t)$ sin ambigüedades. Si $g(t)$ es una función no periódica cuya transformada de Fourier $G_g(\omega)$ está limitada en la banda, como se muestra en la figura 48-34, entonces:

$$g(t) = \int_{-\infty}^{\infty} G_g(\omega) e^{j\omega t} d\omega = \int_{-W}^W G_g(\omega) e^{j\omega t} d\omega \quad (48-59)$$

y $G_g(\omega)$ se desarrolla como una serie de Fourier que converge para $|\omega| < W$:

$$G_g(\omega) = \sum_{n=-\infty}^{\infty} C_n e^{j2\pi n(f/2W)} \quad (48-60)$$

donde:

$$C_n = \frac{1}{2W} \int_{-W}^W G_g(\omega) e^{-j\omega(n/2W)} d\omega \quad (48-61)$$

De este modo, si $g(t)$ se muestrea a intervalos de tiempo regularmente espaciados de $t = -(n/2W)$, cada muestra determinará un coeficiente de Fourier, y si t tiende a infinito, entonces la transformada de Fourier $G_g(\omega)$ está definida. Esto a su vez define la onda de forma temporal, que es simplemente la transformada inversa de Fourier de $G_g(\omega)$. Así, muestrear $g(t)$ a intervalos de tiempo $1/2W$ basta para determinar la función $g(t)$ en todo instante. En la figura 48-35 se ilustra el proceso de muestreo tanto en el dominio del tiempo como en el de la frecuencia. $S(t)$, la función de muestreo, suele ser un tren de pulsos periódicos con ancho d y amplitud constante A . La señal muestreada $m(t)$ está formada por el producto de $S(t)$ y $g(t)$, lo que da por resultado un espectro de frecuencias $G_m(\omega)$ que es la convolución de $G_g(\omega)$ y $G_s(\omega)$. Para el espectro ideal de la señal de información limitado en la banda de la figura 48-34, el espectro de la señal muestreada de $G_g(\omega)$ no se superpondrá si:

$$\frac{1}{T_s} > 2W \quad (48-62)$$

Si la frecuencia de muestreo es dos veces la mayor componente de frecuencia de la señal de información, entonces no habrá superposición de la señal de información en la señal muestreada. Esta frecuencia de muestreo ($2W$) se denomina frecuencia de Nyquist y en teoría es suficiente para recuperar la señal de información original $g(t)$ sin distorsión (Panter).

Para recuperar $g(t)$, la señal muestreada puede hacerse pasar por un filtro pasabajas "ideal" cuya función de transferencia esté definida por:

$$|H(j\omega)|^2 = \begin{cases} 1 & |\omega| < W \\ 0 & |\omega| > W \end{cases} \quad (48-63)$$

En la figura 48-36 se presentan los espectros de frecuencias a la entrada y a la salida del filtro.

Por tanto, el espectro de frecuencias de la salida del filtro pasabajas ideal es el de la señal de información, y

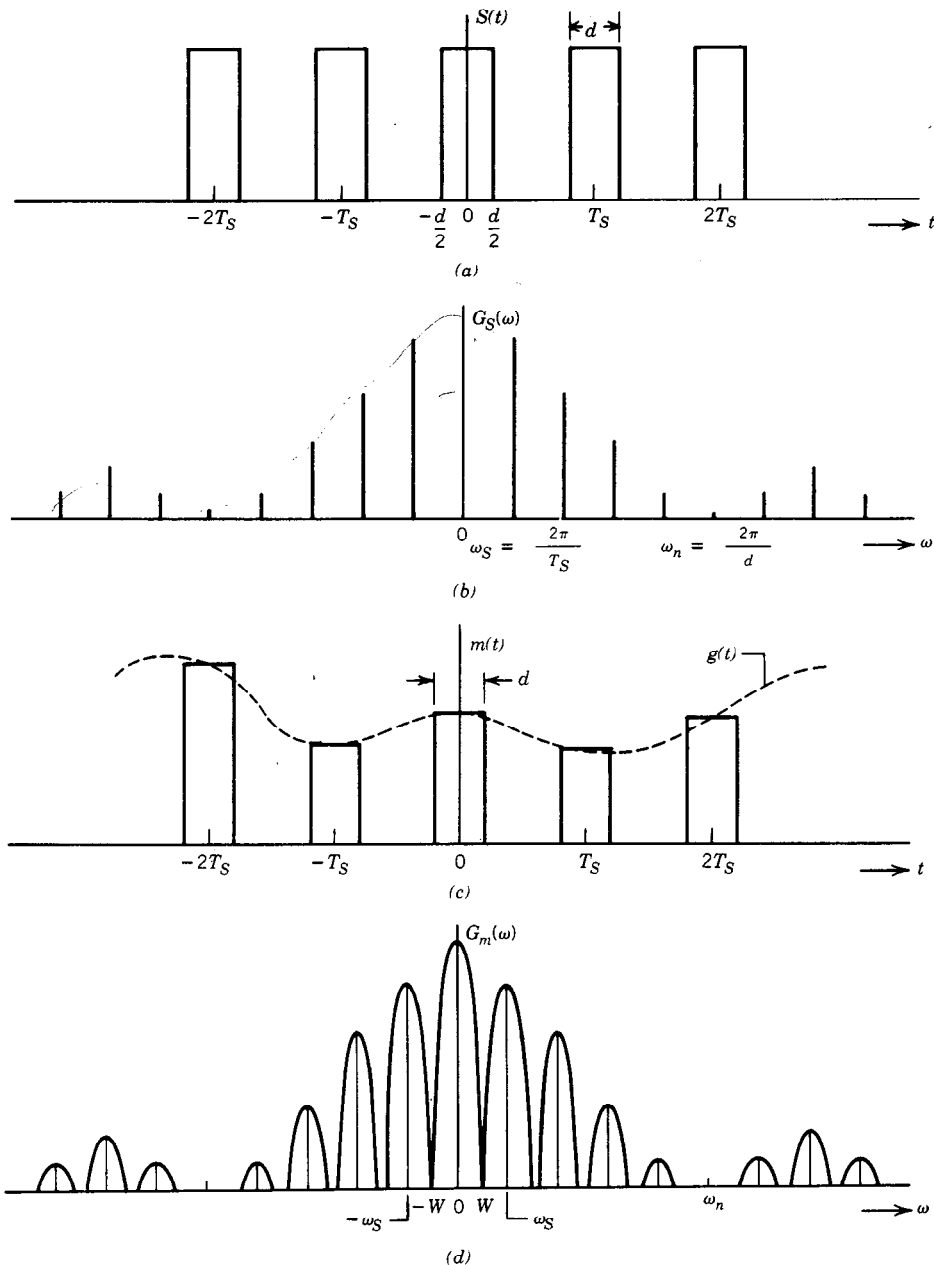


Fig. 48-35. Proceso de muestreo en los dominios del tiempo y de la frecuencia: a) función de muestreo, dominio del tiempo; b) espectro de frecuencia de la señal de muestreo; c) función de la señal muestreada, dominio del tiempo; d) espectro de frecuencia de la señal muestreada.

por ello su función temporal es $g(t)$. Si la frecuencia de muestreo es menor que la frecuencia de Nyquist (es decir, $\omega_S < 2W$), entonces ocurre algo de superposición del espectro muestreado, que introduce un error de aliasing (aparición de una frecuencia inferior) en la señal recuperada, como se muestra en la figura 48-37.

En virtud de que un espectro idealmente limitado en banda sólo existe en teoría, siempre habrá algo de alia-

sing y distorsión resultantes. Sin embargo, la distorsión puede reducirse a niveles aceptables incrementando ω_S , la frecuencia de muestreo.

Modulación de pulsos en amplitud (PAM)

Si la señal muestreada $m(t)$ definida en la figura 48-35 se transmite directamente, el tren de pulsos resultante es el de una señal modulada de pulsos en amplitud

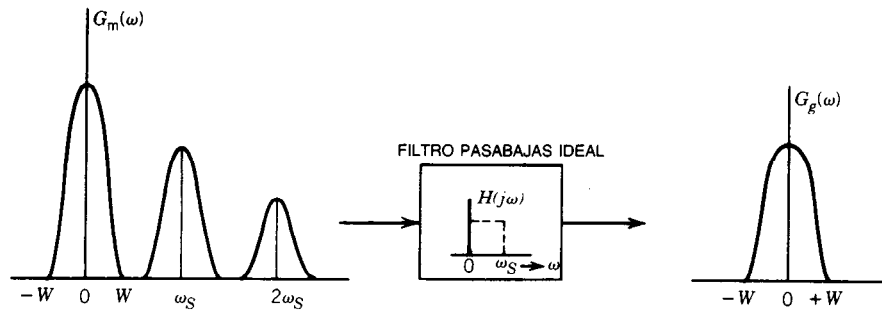


Fig. 48-36. Espectros de frecuencia de la señal muestreada y recuperada después de pasar por un filtrado pasabajas.

(PAM, de *pulse amplitude-modulation*). Si la potencia promedio del tren de pulsos sin modular se denota por P_p , entonces, cuando se introduce la modulación, como se muestra en la figura 48-35, la potencia de la señal es $m_p^2 \bar{g}^2(t) P_p$, donde m_p es el índice de modulación o factor de modulación del tren de pulsos. Entonces, puede demostrarse que la razón de salida de señal sobre ruido, suponiendo un filtro pasabajas ideal, es (Panter):

$$(S/N)_{o.PAM} = \frac{m_p^2 \bar{g}^2(t) (\tau/T_S) P_p}{N_0 W} \quad (48-64)$$

donde W = mayor frecuencia de la señal de información limitada en la banda

$N_0/2$ = densidad espectral bilateral de ruido

τ = ancho de pulso de la señal de muestreo

T_S = frecuencia de pulsos de la señal de muestreo.

Modulación de pulsos en duración

Otra técnica de modulación de pulsos consiste en variar el tiempo de ocurrencia del flanco ascendente o descendente del pulso, lo que da por resultado la variación del ancho o la duración del pulso. En la figura 48-38 se ilustra una señal modulada de pulsos en duración y la función temporal de información o modulación.

Si los pulsos más largo y más corto se denotan por T_L y T_X respectivamente, y la información del mensaje se considera como si residiera directamente en la duración de los pulsos, entonces la duración media de los pulsos \bar{T} representará una muestra de amplitud de una señal con información cero y es:

$$\bar{T} = \frac{1}{2}(T_L + T_X) \quad (48-65)$$

La excursión positiva y negativa de escala completa de la señal de información se representa entonces por medio de la duración temporal:

$$(\bar{T} - T_X) = (T_L - \bar{T}) = \frac{1}{2}(T_L - T_X) \quad (48-66)$$

Así, para la desmodulación de la medición temporal directa de la señal PDM, la potencia de la señal de salida, suponiendo una señal de información continua $g(t)$ de amplitud cresta unitaria, es:

$$S_{o.PDM} = \frac{1}{4}(T_L - T_X)^2 \bar{g}^2(t) \quad (48-67)$$

El ruido de salida en un sistema PDM se manifiesta como fluctuaciones en el flanco descendente del pulso. Dado que los flancos ascendentes de los pulsos no están modulados y ocurren a frecuencia regular, pueden utilizarse para reducir una referencia de sincronización sin ruido. En la figura 48-49 se muestra la caracteri-

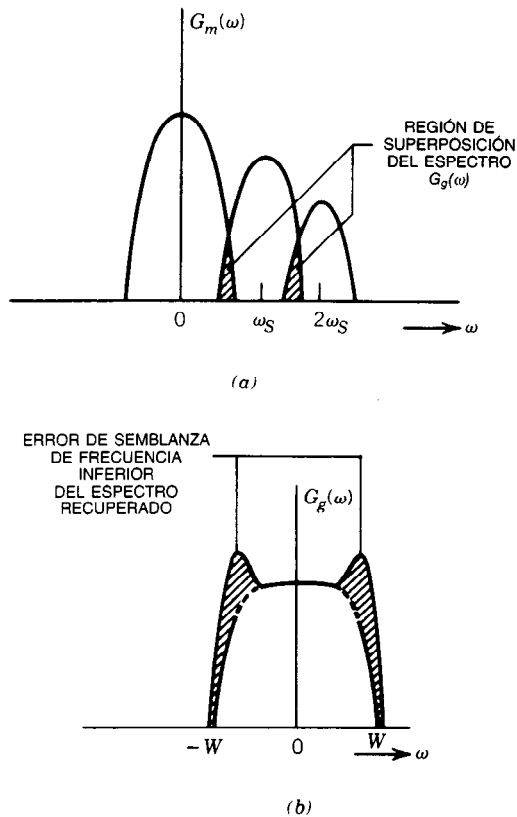


Fig. 48-37. Aliasing debido a una frecuencia de muestreo insuficiente ($\omega_s < 2W$): a) espectro de la señal muestreada; b) espectro de la señal recuperada.

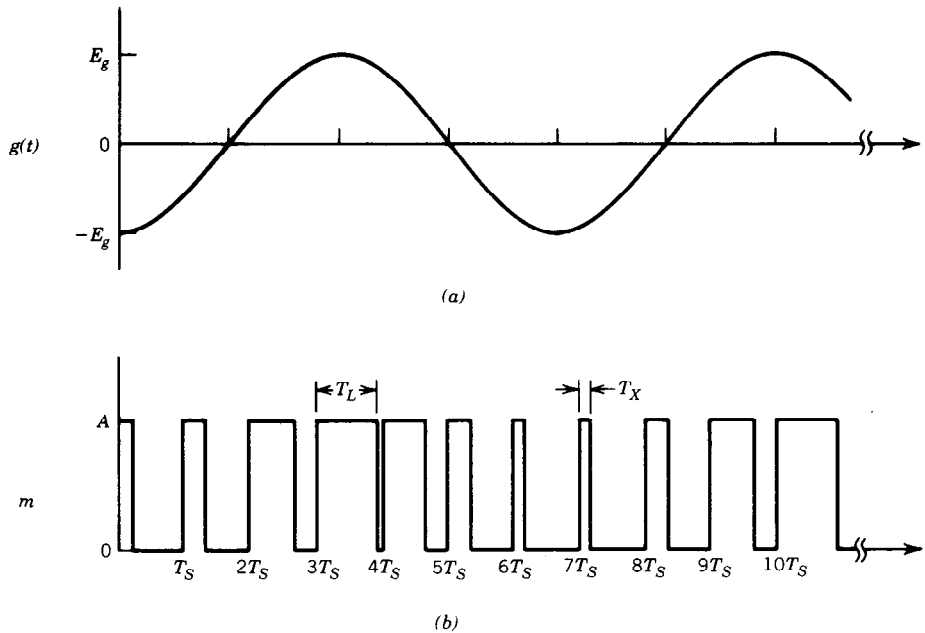


Fig. 48-38. Función del tiempo de una señal modulada en duración de pulsos; a) señal de información, $g(t) = E_g \sin \omega_g t$; b) señal modulada en duración de pulsos, $m_{PDM}(t)$.

zación de ruido del sistema PDM. La potencia de salida del ruido del proceso de desmodulación temporal es:

$$n_t = \frac{n_n}{\lambda^2} \quad (48-68)$$

donde $\sqrt{n_n}$ es la desviación estándar del ruido de la banda base (voltaje eficaz o rms de ruido) y λ es la pendiente del flanco descendente del pulso.

La razón de salida de señal sobre ruido puede expresarse como (Downing):

$$(S/N)_{\text{sal. PDM}} = \frac{\lambda^2 (T_L - T_X)^2 \overline{g^2(t)}}{4n_n} \quad (48-69)$$

Obsérvese que la razón de salida de señal sobre ruido depende de la relación λ^2/n_n , por lo que es necesario encontrar una solución intermedia entre estas cantidades con respecto al filtrado del ruido de la banda base. Dado que la máxima pendiente de la respuesta escalón de un filtro pasabajas de un tipo dado es directamente proporcional a su ancho de banda, la presencia de λ en el numerador es un argumento a favor del máximo ancho de banda posible del filtro de banda base; sin embargo, a medida que se incrementa el ancho de banda base de este filtro, la potencia del ruido aumenta de forma monótona, y esta cantidad en el denominador favorece el filtro de banda base más estrecho posible. En los sistemas AM, donde la densidad espectral de potencia del ruido es uniforme, la potencia del ruido de la banda base aumenta con la primera potencia del ancho de banda del filtro de banda

base. Así, el mejor rendimiento del sistema debería obtenerse si no se utilizara filtrado de banda base.

Modulación de pulsos en posición

Si la señal modulada en duración de pulsos es diferenciada y se genera otro tren de pulsos de amplitud y duración constantes, de modo que coincida con los picos negativos de la onda diferenciada, entonces se obtiene una señal modulada de pulsos en posición (PPM, de *pulse position-modulated*). En la figura 48-40 se ilustra el proceso de PPM por diferenciación de una señal PDM.

Un sistema PPM puede considerarse análogo a un sistema modulado en ángulo, ya que la información está contenida en la posición de los flancos ascendentes del pulso con respecto a la posición media o promedio de los flancos. En la figura 48-41 se ilustran los efectos del ruido del sistema en la señal PPM recuperada.

Si T_o es el máximo desnivel respecto al tiempo medio de ocurrencia del pulso en ambas direcciones, positiva y negativa, y \sqrt{nt} es el error temporal o fluctuación debido al ruido del sistema con desviación estándar $\sqrt{n_n}$, entonces la razón de salida de señal sobre ruido es una función de la pendiente del flanco ascendente λ del tren de pulsos, donde:

$$\lambda = \frac{\sqrt{nt}}{\sqrt{n_n}} \quad (48-70)$$

o bien:

$$nt = \frac{n_n}{\lambda^2} \quad (48-71)$$

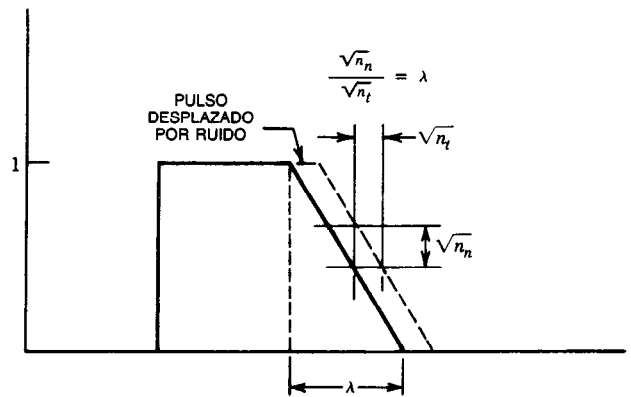


Fig. 48-39. Variación en el ancho de los pulsos debido al ruido de banda ancha del sistema con desviación estándar $\sqrt{n_n}$.

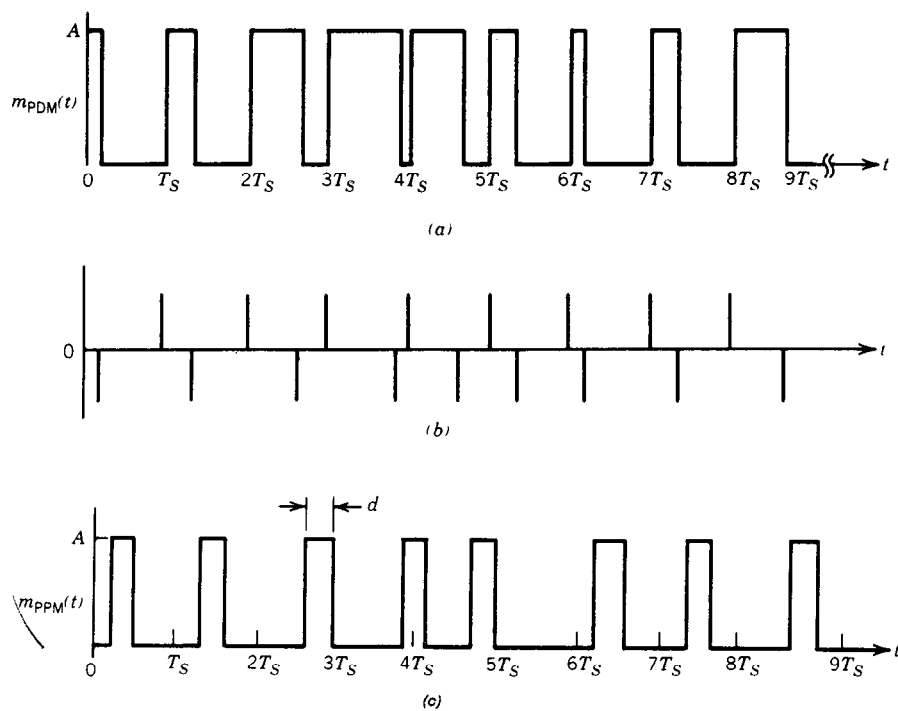


Fig. 48-40. Modulación en la posición de los pulsos (PPM); a) señal modulada en duración de pulsos (PDM), función del tiempo; b) señal PDM diferenciada, función del tiempo; c) señal PPM, función del tiempo.

Suponiendo que se utiliza desmodulación temporal directa, la potencia de la señal de salida es:

$$S_{p,PPM} = kT_o^2 \overline{g^2(t)} \quad (48-72)$$

Así, la razón de salida de señal sobre ruido es (Panter):

$$(S/N)_{sal,PPM} = \frac{T_o^2 \overline{g^2(t)} \lambda^2}{n_n} \quad (48-73)$$

Una vez más, es evidente que el incremento en la pendiente λ de los flancos ascendentes del pulso mejorará la razón de salida de señal sobre ruido, siempre que el espectro de potencia del ruido de salida sea plano.

Por tanto, un criterio de ancho de banda amplio semejante al de los sistemas de modulación PDM producirá el rendimiento óptimo de la SNR para la señal de información recuperada.

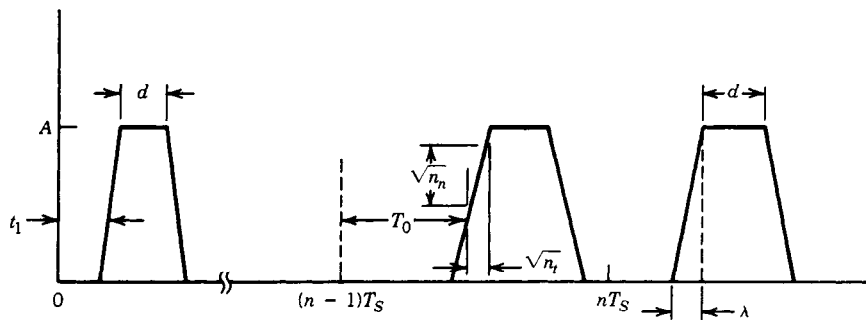


Fig. 48-41. Variación en la posición de los pulsos debido al ruido de banda ancha del sistema con desviación estándar $\sqrt{n_n}$.

Modulación de pulsos en código

En los métodos previamente considerados de modulación de pulsos se representaban señales mediante el muestreo de funciones analógicas. También es costumbre representar señales de información continua en forma digital codificada. En este proceso de digitalización, la señal de información $g(t)$ debe muestrearse en el tiempo, y cada muestra debe cuantizarse en amplitud.

En la figura 48-42 se ilustra una señal de información muestreada cuantizada en amplitud.

El intervalo de cuantización Q depende de la resolución requerida y del nivel de ruido de cuantización que los canales de información pueden tolerar. En la figura 48-42, en el instante t_1 el sistema debería transmitir $\gamma_q = 3$; en el instante t_2 , $\gamma_q = -1$; en el instante t_3 , $\gamma_q = -2$, etcétera. De este modo, la información realmente transmitida es el símbolo que representa el número de cuantización individual, en vez de la amplitud muestreada exacta de la señal de información. El número de intervalos de cuantización (QI) está relacionado con la resolución deseada y con la amplitud de cresta a cresta de la señal de información.

Si el intervalo de cresta a cresta de $g(t)$ es E_g y la resolución deseada es R , entonces el tamaño del intervalo de cuantización es $Q = RE_g$. El número de intervalos de cuantización requerido sería:

$$N_q = \frac{E_g}{Q} = \frac{1}{R} \quad (48-74)$$

Ruido de cuantización. En la figura 48-43 se muestran los efectos del ruido resultante del proceso de cuantización. En el instante t_1 el transmisor transmitirá el símbolo que representa $\gamma_q = 3$, y en el receptor a la muestra recuperada de $g(t_1)$ se le asignará el valor correspondiente al centro del número de cuantización $\gamma_q = 3$. Por tanto, existe un error (q) en la amplitud de la señal de información recuperada. q es una variable aleatoria que representa la diferencia entre $g(t)$ y el valor medio del número de cuantización.

Si se supone que la variable aleatoria q está uniformemente distribuida en el intervalo de cuantización Q , entonces la distribución de probabilidad del error de cuantización es uniforme, como se muestra en la figura 48-44.

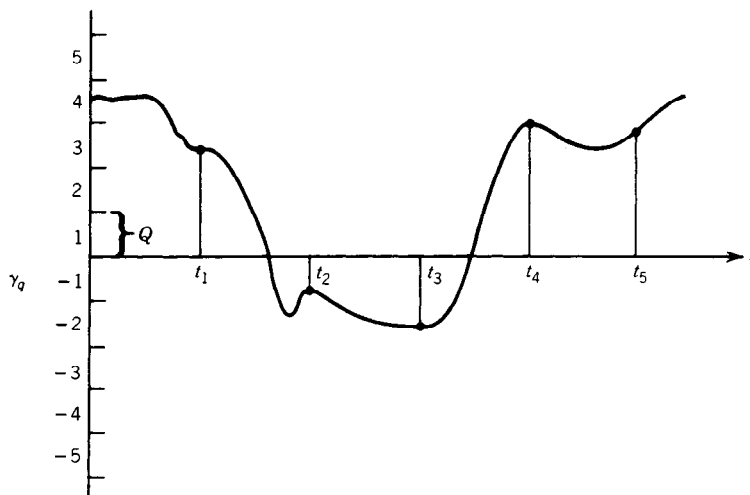


Fig. 48-42. Cuantización de una onda de forma continua.

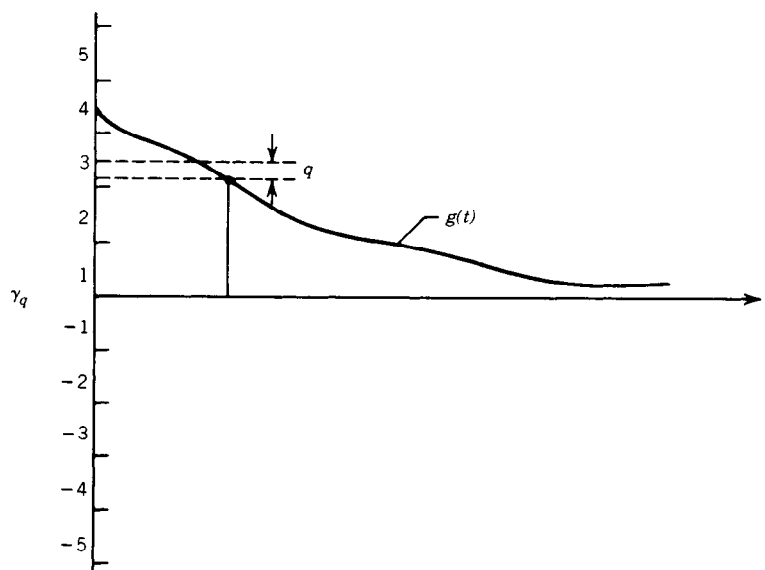


Fig. 48-43. Ruido de cuantización.

La potencia de ruido que se obtiene de este proceso de error aleatorio es (Downing):

$$P_Q = \frac{Q^2}{12} = \frac{RE_g^2}{12} = \frac{E_g^2}{12\gamma_q^2} \quad (48-75)$$

Si la potencia de la señal de información está dada por $g^2(t)$, entonces la razón de señal sobre ruido de cuantización pasa a ser (Downing):

$$(S/N)_q = \frac{12\overline{g^2(t)\gamma_q^2}}{E_g^2} \quad (48-76)$$

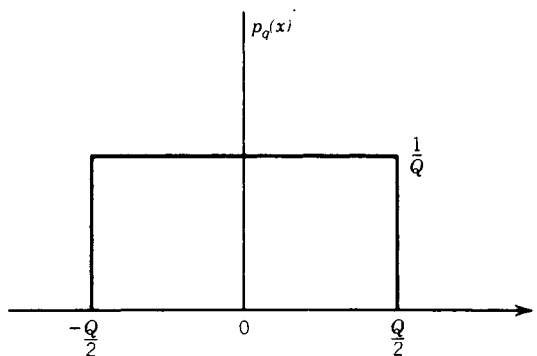
Digitalización. Una vez que se han muestreado y cuantizado las amplitudes de la señal, es posible codificar digitalmente los intervalos de cuantización, y transmitir en el instante muestreado el código digital que representa el intervalo de cuantización idóneo. Por ejemplo, si $N_q = 8$, entonces la figura 48-45a ilustra este proceso de codificación binaria. Cada uno y cero del número binario que representa cada intervalo de cuantización se denomina bit, y cada grupo de bits es una palabra. Con palabras de N bits es posible distinguir entre 2^N intervalos de cuantización. En la tabla 48-6 se muestran los tamaños de palabra en bits, el número de intervalos de cuantización y la resolución en porcentaje.

Los símbolos binarios transmitidos forman una corriente continua de bits, y la frecuencia (velocidad) de esta corriente continua debe ser lo suficientemente alta para conformarse a la frecuencia inicial de digitalización (muestreo) multiplicada por las longitudes de la palabra binaria (número de bits por palabra). En la figura 48-45b se ilustra la onda de forma temporal binaria final transmitida, suponiendo un formato de datos binario sin regreso a cero.

El sistema en que se cuantiza el mensaje tanto en tiempo como en amplitud y en que se utilizan los símbolos binarios para transmitir la muestra de información se denomina sistema de modulación de pulsos en código o de modulación por codificación de pulsos (PCM, de *pulse code modulation*). La mínima velocidad de datos PCM requerida es igual al tamaño de la palabra dividido entre la frecuencia de muestreo. Por consiguiente, si las muestras de amplitud se toman a intervalos espaciados regularmente por 1 ms y el tamaño de la palabra PCM es de ocho bits, se necesitaría una velocidad de transmisión de datos PCM mínima de 8 kb/s (kilobits por segundo).

Modulación en delta

La modulación en delta es un tipo de sistema de cuantización predictivo equivalente a un sistema de modulación de pulsos de código diferencial de un dígito. Estos

Fig. 48-44. Distribución de probabilidad del error de cuantización (q).

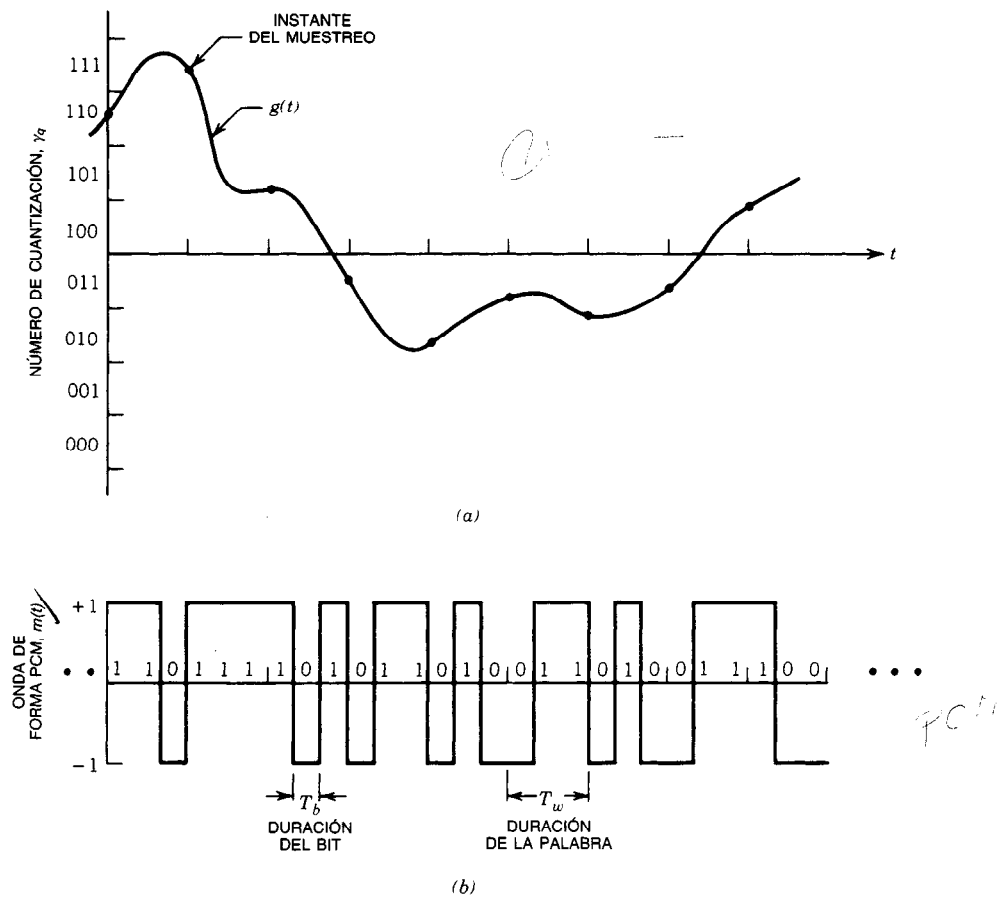


Fig. 48-45. Ejemplo de cuantización [modulación por codificación de pulsos (PCM)]: a) proceso de codificación binario; b) onda de forma temporal binaria transmitida, suponiendo un formato de datos binario sin regreso a cero (NRZ).

Estos sistemas se basan en la transmisión de la diferencia cuantizada entre valores muestrales consecutivos, más que en las muestras en sí. Tanto los moduladores como los desmoduladores hacen una estimación o predicción del valor de la señal de entrada con base en la sucesión previamente transmitida.

Existen dos tipos de modulación en delta: la lineal y la adaptativa. En la modulación lineal en delta, se predice el valor de la señal de entrada en cada instante de muestreo como una función lineal particular de los valores anteriores de la señal. El efecto es hacer la diferencia entre cualesquiera dos valores predictores consecutivos igual al nivel de cuantización. En la modulación adaptativa en delta, el valor de la señal de entrada en cada instante de muestreo se predice como una función no lineal de los valores anteriores de la señal. La introducción de la predicción no lineal en la modulación en delta constituye una forma útil de ampliar el intervalo en el cual el sistema produce su rendimiento óptimo.

En la figura 48-46 se muestra un diagrama de bloques de un modulador en delta. La entrada es una

Tabla 48-6. Tamaños de palabra de modulación por codificación de pulsos contra resolución requerida y número de intervalos de cuantización

Tamaño de palabra N (núm. de bits)	Número de intervalos de cuantización $\gamma_q = 2^N$	Resolución $R = 1/\gamma_q(\%)$
1	2	50
2	4	25
3	8	12.5
4	16	6.25
5	32	3.125
6	64	1.563
7	128	0.78125
8	256	0.391
9	512	0.195
10	1 024	0.098

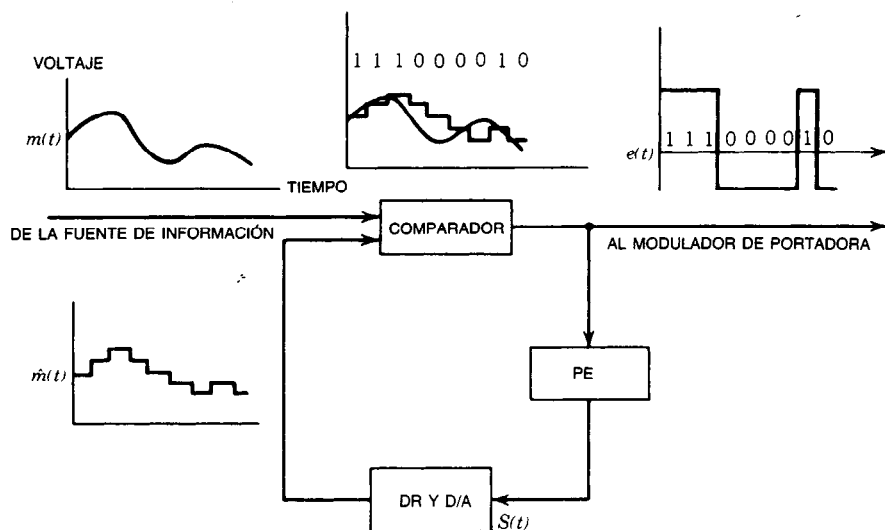


Fig. 48-46. Modulación en delta, diagrama funcional. PE = procesador del tamaño del escalón; DR = dispositivo de retroalimentación.

señal de información analógica cuyo valor instantáneo se digitaliza por técnicas de modulación en delta. En el funcionamiento del modulador, la salida analógica del dispositivo de retroalimentación se compara con la señal analógica de entrada. El resultado de esta comparación, que es una señal digital, hace que el dispositivo de retroalimentación incremente o reduzca la magnitud de su salida para aproximarla a la de la entrada. La velocidad con que el modulador cambia su magnitud está controlada por una señal de reloj. La corriente de bits fluye a la velocidad del reloj.

El funcionamiento del desmodulador es idéntico, ya que la desmodulación ocurre a la salida del dispositivo de retroalimentación. Así, la salida del desmodulador es una reproducción analógica de la señal de entrada predicha por el modulador. Si en el proceso de transmisión ocurren errores en los bits, la corriente de bits del desmodulador no es idéntica a la de la salida del modulador, por lo que la salida analógica del desmodulador no reproducirá con precisión la señal de entrada del modulador. Sin embargo, en la generalidad de los casos la modulación en delta tiene una elevada tolerancia a errores en los bits y, por tanto, puede esperarse una leve degradación en presencia de ruido. En el caso en que la señal de información es el habla, el factor de inteligibilidad del canal del habla no se degrada de forma significativa hasta que la proporción de errores es del orden de un error en 10 bits, o 1×10^{-1} . Esto está muy por debajo del intervalo utilizable para la mayor parte de los sistemas digitales tipo PCM. En la figura 48-47 se muestra una gráfica de inteligibilidad de palabra (basada en palabras fonéticamente equilibradas) como función de la proporción de errores en los bits para el algoritmo de modulación adaptativa en delta (fig. 48-48) utilizado en los canales de audio de los transbordadores espaciales. Estos canales funcionan a una velocidad de reloj de modulación en delta de 32 kb/s (Seyl y colaboradores).

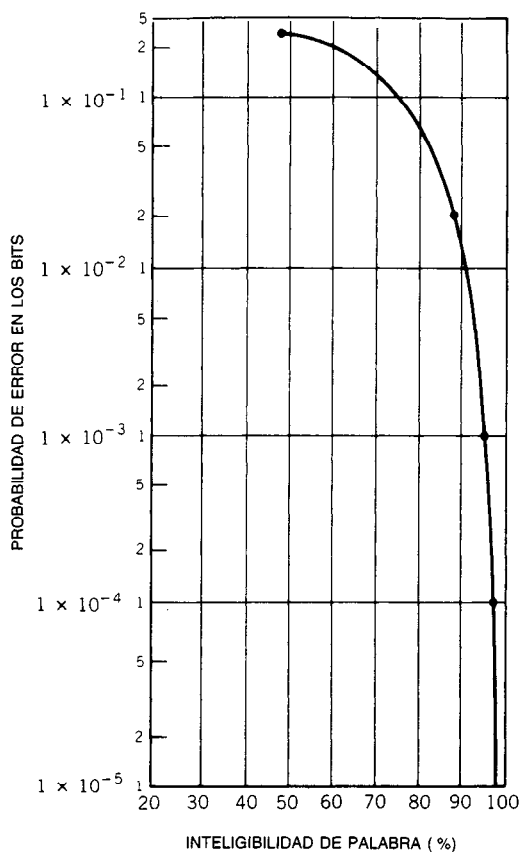


Fig. 48-47. Inteligibilidad de la palabra hablada en función de la probabilidad de errores en los bits para el sistema de modulación en delta de los transbordadores espaciales. Tomado de J. W. Seyl y colaboradores, con permiso.

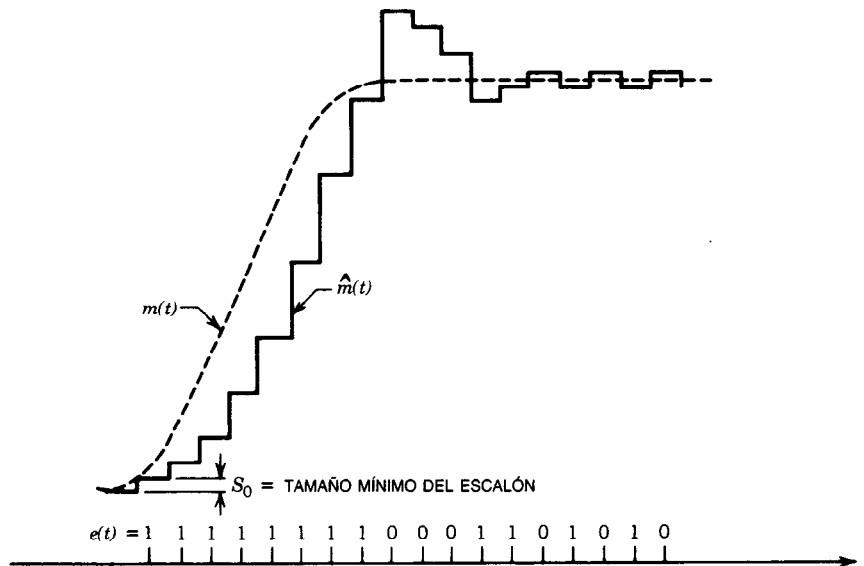


Fig. 48-48. Algoritmo de modulación adaptativa en delta, utilizado para enlaces digitales del habla de los transbordadores espaciales.

$$S(t+1) = \begin{cases} [|S(t)| + S_0]v(t) & \text{cuando } e(t) = e(t-1) \text{ y } |S(t)| < 8S_0 \\ |S(t)|v(t) & \text{cuando } e(t) = e(t-1) \text{ y } |S(t)| = 8S_0 \\ |S_0|v(t) & \text{cuando } e(t) \neq e(t-1) \end{cases}$$

$$v(t) = \begin{cases} +1, & e(t) = 1 \\ -1, & e(t) = 0 \end{cases}$$

Multicanalización y desmulticanalización

Para transmitir de forma simultánea más de una señal de información por el mismo canal es necesario utilizar algún método para combinar o multicanalizar las señales. En cualquier canal de comunicación en el que sea posible la multicanalización existen dos dominios: el dominio del tiempo, en el que ocurre la multicanalización por división de tiempo (TDM, de *time division multiplexing*), y el dominio de la frecuencia, en el que se da la multicanalización por división de frecuencia (FDM, de *frequency division multiplexing*). Además, se han inventado formas especiales de multicanalización, como la multicanalización por división de códigos, en la cual cada señal de información se identifica mediante un código pseudoaleatorio único, y la multicanalización por portadora en cuadratura, en la que se utiliza una señal portadora con componentes de cuadratura. En la siguiente sección se analizará cada una de estas formas de multicanalización.

Multicanalización por división de frecuencia (FDM).

Como su nombre implica, en el sistema FDM se combinan múltiples canales de información mediante la combinación de señales subportadoras de diferentes frecuencias en el transmisor y filtrado de las señales recuperadas a fin de seleccionar el canal de información individual en el receptor. En la figura 48-49 se ilustra el sistema FDM.

Si los espectros de la señal de información están idealmente limitados en la banda y se utiliza modulación de doble banda lateral con supresión de portadora como método de modulación de la subportadora, entonces las frecuencias de ésta pueden elegirse de modo que se evite la superposición en el espectro de la señal de información, como se observa en la figura 48-50.

Suponiendo un formato de modulación de subportadora DSBSC, el espectro de la señal de banda base consta de las frecuencias individuales de la subportadora con el espectro correcto de la señal de información trasladado a cada frecuencia de la subportadora. La separación entre las componentes espectrales superiores de la banda lateral del canal uno y las inferiores de la banda lateral del canal dos suele denominarse banda de guarda o de seguridad. Por consiguiente, para que no exista superposición del espectro de la información, la separación requerida entre las frecuencias de la subportadora es $2W$ más la banda de guarda o de seguridad. El espectro combinado de la señal FDM a su vez modula la señal portadora. La modulación de ésta puede ser AM, PM o FM, y para las dos primeras, el espectro de la señal detectada en la banda base se vería como se muestra en la figura 48-51.

En virtud de que el espectro del ruido a la salida del receptor es plano para todas las frecuencias, el empleo de factores iguales de modulación de amplitud o fase para cada canal de subportadora daría por resultado

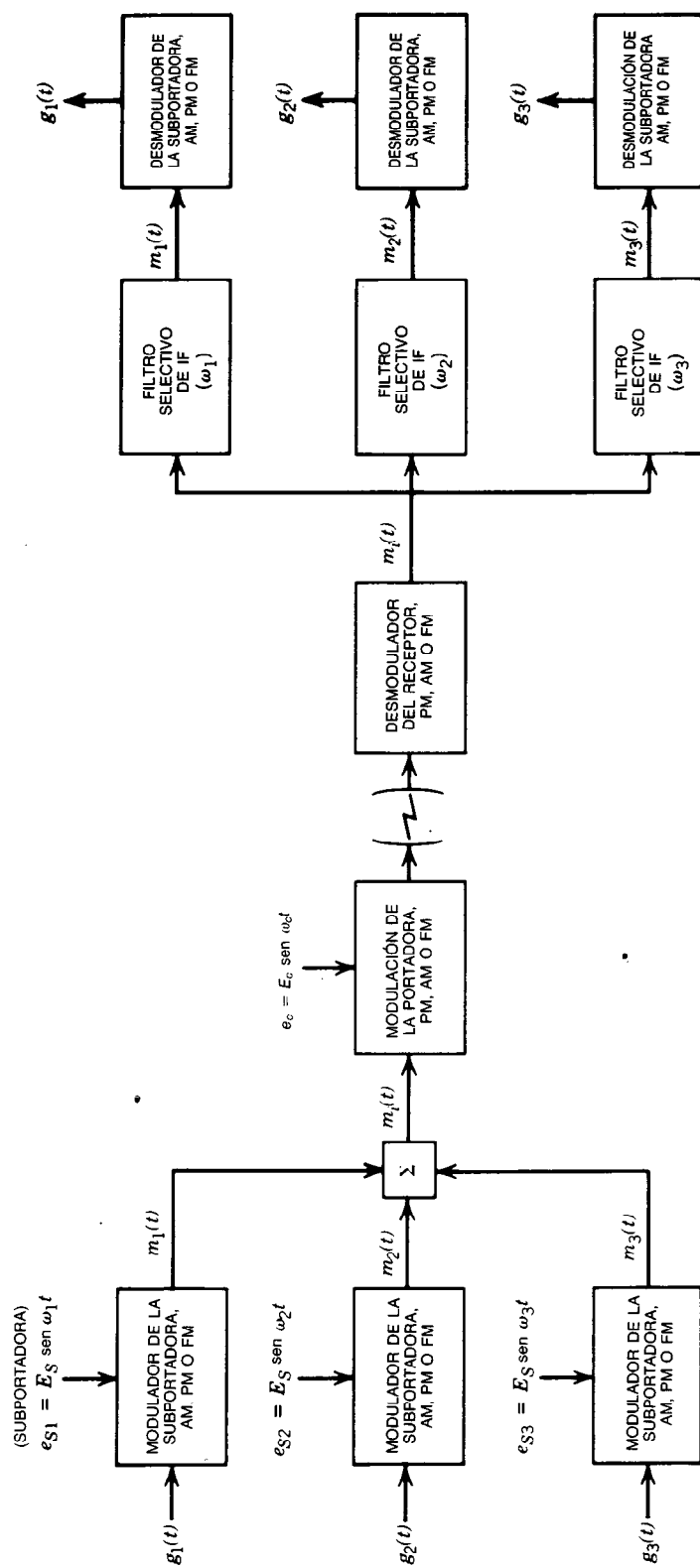


Fig. 48-49. Sistema para transmisión de señales empleando multicanalización por división de frecuencia, diagrama funcional.

el mismo rendimiento de señal sobre ruido para cada canal, suponiendo iguales anchos de banda de información. Sin embargo, esto no ocurre para la modulación en frecuencia de la portadora debido a la naturaleza parabólica de la amplitud del ruido de salida del receptor de FM. En la figura 48-52 se ilustra una señal de salida FDM de un canal FM.

La señal moduladora compuesta es $m(t)$:

$$m(t) = \sum_{i=1}^V K_i m_i(t) \quad (48-77)$$

donde K_i representa el factor de ponderación individual del canal de la subportadora, que determina su índice de modulación en la señal compuesta. Para asegurar que el modulador de frecuencia de la portadora no esté sobremodulado, es necesario elegir la desviación de la frecuencia de cresta con base en estos factores de ponderación y en la probabilidad de que la señal compuesta exceda un valor específico. Si están presentes muchas subportadoras y no se encuentran relacionadas coherentemente en frecuencia o fase, entonces una hipótesis razonable es una distribución de probabilidad gaussiana, y Δf puede establecerse utilizando el valor 3σ como la desviación pico de la señal compuesta FDM (Downing):

$$\Delta f^2 = 9 \sum_{i=1}^V K_i^2 \overline{m_i^2(t)} \quad (48-78)$$

Si $\overline{m_i^2(t)}$ está normalizado a la unidad, entonces:

$$\Delta f^2 = 9 \sum_{i=1}^V K_i^2$$

con probabilidad del 99.7%.

Suponiendo una densidad espectral de potencia para el ruido de salida de la banda base, como se muestra en la figura 48-52, la SNR de salida para el i -ésimo canal sería:

$$(S/N)_{o,FDM}^{(i)} = \frac{E_c^2 K_i^2}{2K^2 f_i^2 B_m} \quad (48-79)$$

Por consiguiente, para un rendimiento idéntico SNR (razón señal-ruido) de canales de información, el factor de ponderación K_i debe estar dado a la frecuencia de la subportadora del canal, f_i . La proporcionalidad depende del número de canales V , del ancho de banda de información B_m y de la restricción de la modulación de cresta de la portadora.

Así, cuantitativamente para un mismo rendimiento de SNR de canales se requiere que los canales de fre-

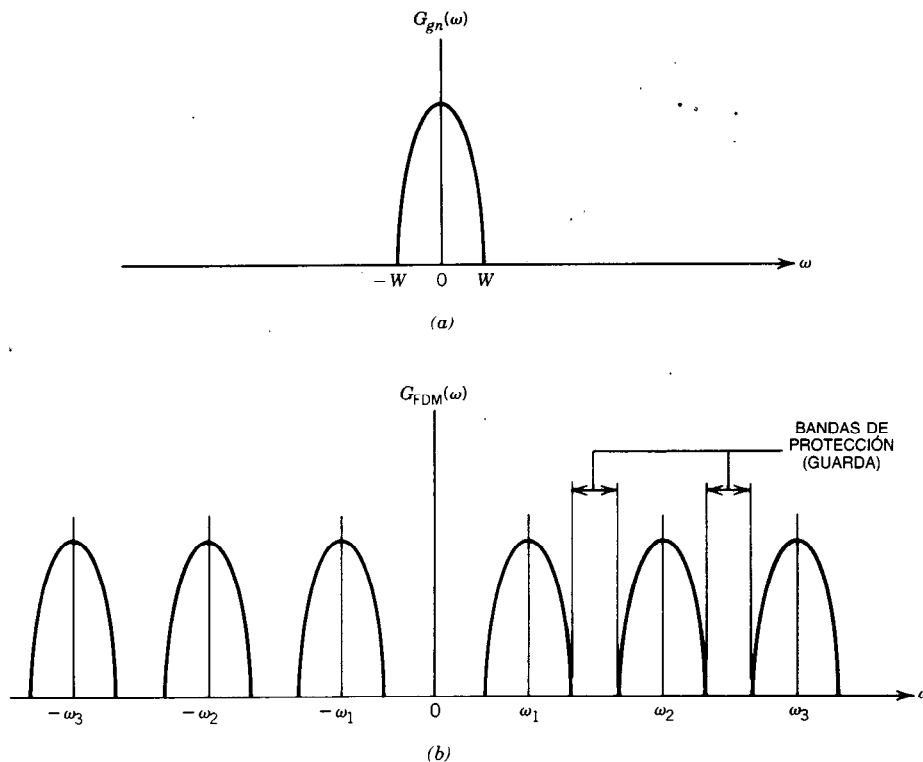


Fig. 48-55. Formato típico de cuadros de datos para telemetría por modulación por codificación de pulsos (PCM), para datos de multicanalización por división en el tiempo (TDM). Patrón de sincronización: 76571440₈. 128 kilobits por segundo, 8 bits por palabra, 160 palabras por cuadro secundario, 100 cuadros secundarios por cuadro principal, y 1 cuadro principal por segundo.

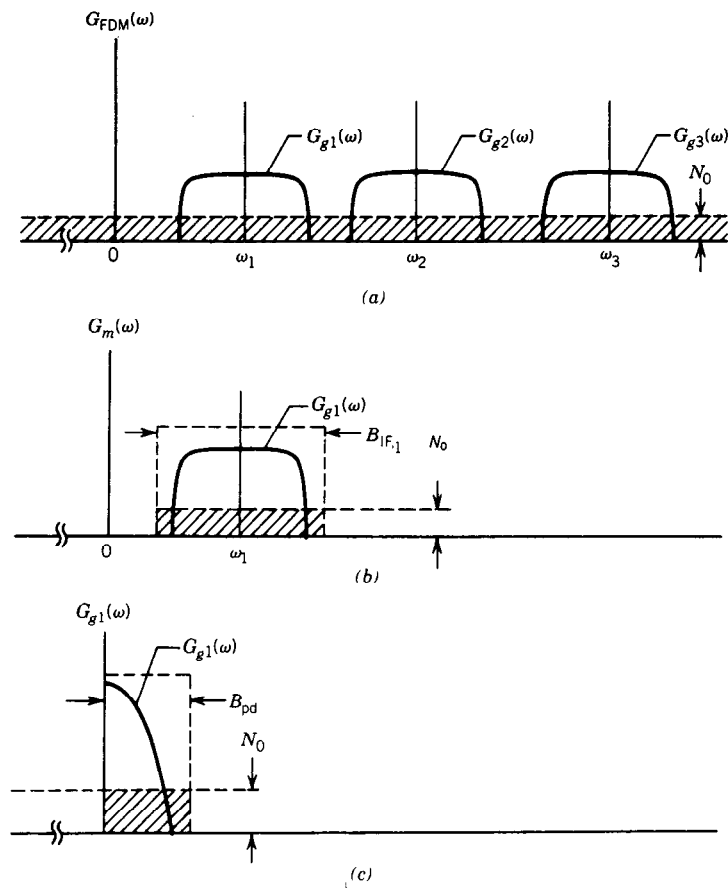


Fig. 48-51. Espectros típicos de la señal detectada mediante multicanalización por división de frecuencia (FDM) para un canal modulado en frecuencia (AM) o fase (PM): a) espectro detectado en la banda base más el ruido del sistema; b) espectro de la señal de salida del filtro pasabanda selectivo (ω_1) más el ruido del sistema; c) espectro de la señal posdetectada para el canal 1 más el ruido.

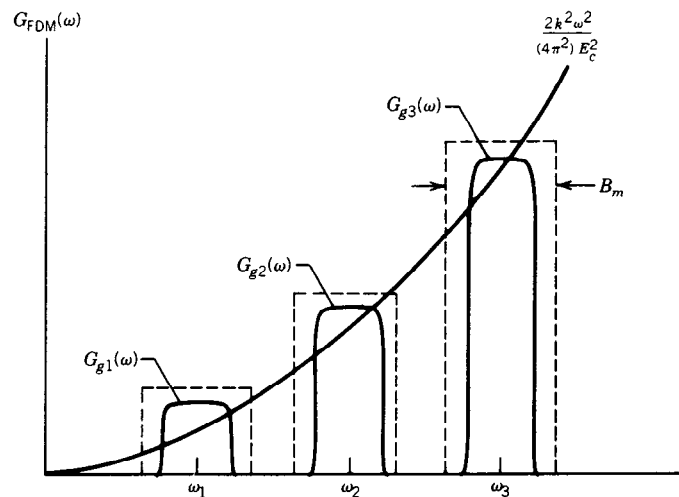


Fig. 48-52. Señal de salida multicanalizada por división de frecuencia (FDM) para un canal de comunicación modulado en frecuencia.

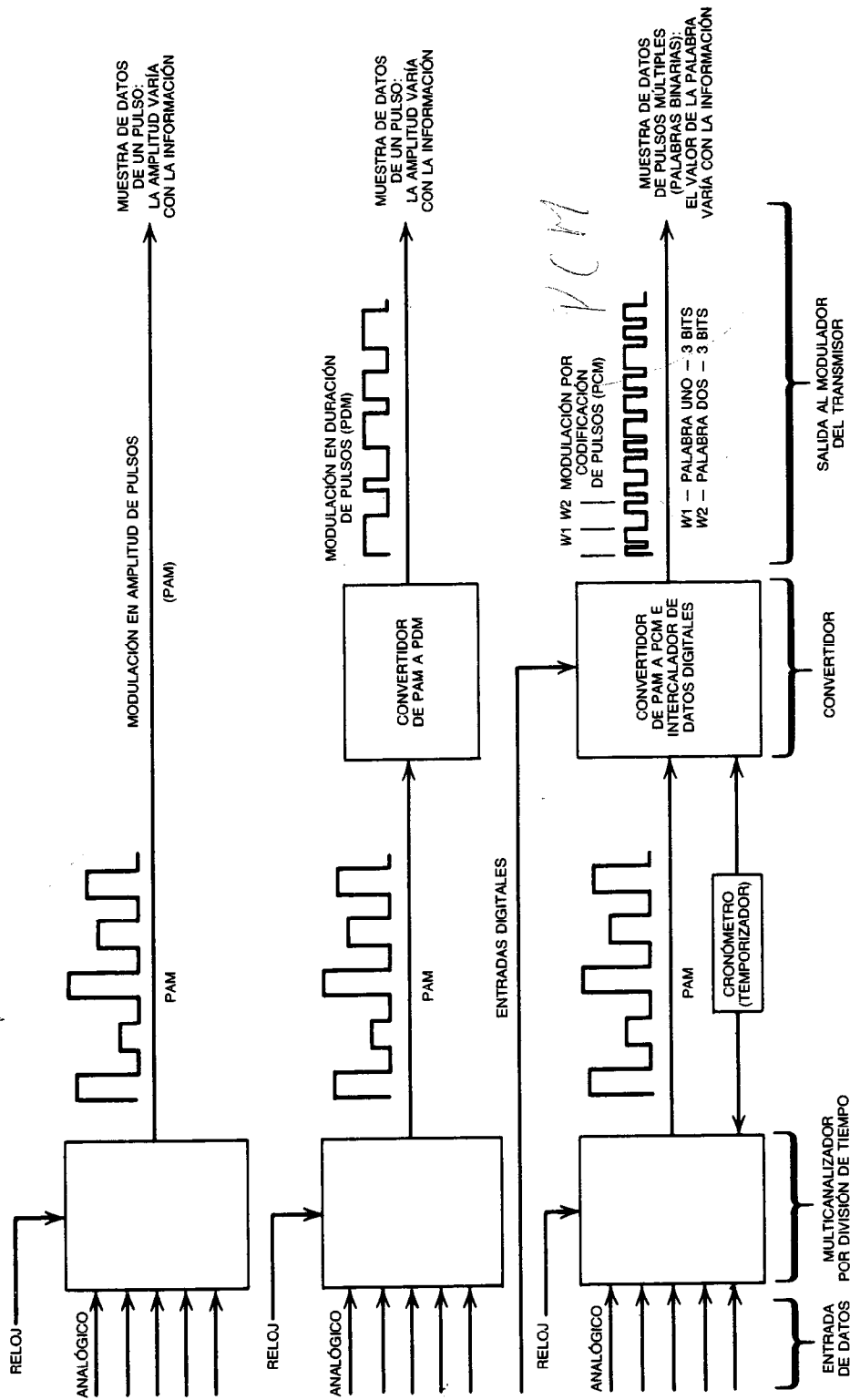


Fig. 48-53. Multicanalización por división de tiempo (TDM), diagrama funcional.

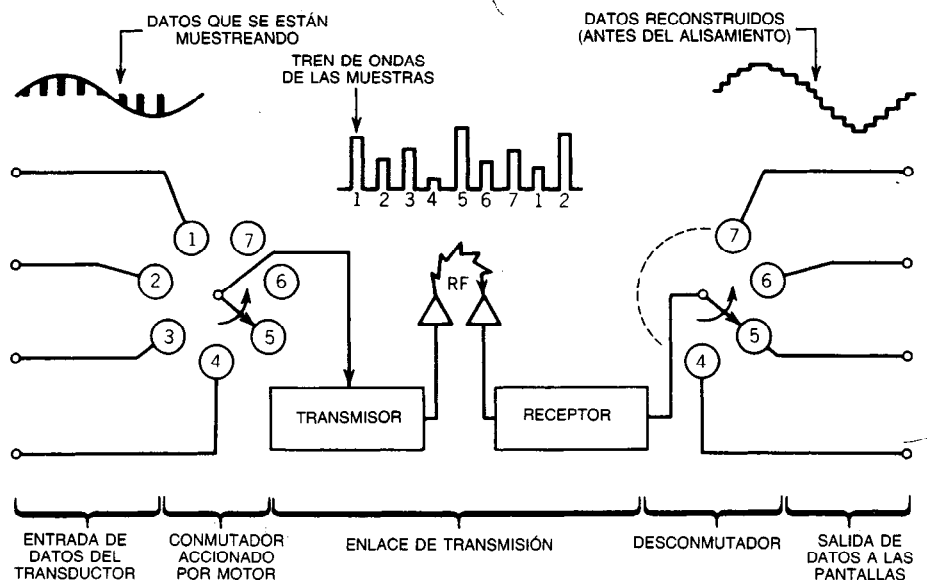


Fig. 48-54. Multicanalización por división de tiempo, en la que se utiliza un interruptor giratorio como conmutador.

cuencias superiores desvíen la portadora más que los canales de frecuencias inferiores. Por tanto, suele utilizarse alguna forma de preacentuación o preénfasis para ajustar los factores de ponderación en los sistemas de multicanalización FDM-FM.

Multicanalización por división de tiempo. En este tipo de multicanalización se combinan señales de información en el dominio del tiempo. Como ya se analizó para la modulación de pulsos, las señales de información pueden muestrearse en el tiempo y reconstruirse en el receptor a partir de las muestras temporales transmitidas. Si varias señales de información se muestrean a diferentes intervalos de tiempo y las señales muestreadas resultantes se combinan por un método sincrónico, entonces la forma de onda temporal se denomina señal multicanalizada por división de tiempo (TDM). En la figura 48-53 se ilustra la señalización TDM utilizando técnicas PAM, PDM y PCM.

En la figura 48-54 se ilustra el muestreo utilizando un interruptor giratorio como conmutador y desconmutador. El proceso requiere que el conmutador y el desconmutador giren sincrónicamente en el tiempo a fin de asegurar que la muestra correcta de la señal de información aparezca en la terminal de salida deseada. Un método de sincronización de uso común es la reconstrucción de la señal del reloj mediante el sistema receptor, que puede utilizarse para asegurar que el desconmutador gire a la misma velocidad que el conmutador. Sin embargo, además de la coordinación temporal básica establecida por el reloj, debe contarse con alguna forma de identificación del canal, a fin de establecer la localización de un canal de referencia. Esto puede lograrse dedicando un canal del sistema a una señal definitoria o impulso marcador. Cuando se

utiliza la modulación de pulsos en código de los formatos TDM, el proceso de conmutación y desconmutación se verifica a través de la formación de cuadros de datos de información. Cada cuadro contiene palabras de datos que se localizan en segmentos de tiempo con respecto a una palabra de sincronización que identifica el inicio del cuadro de datos. Además, los cuadros de datos subconmutados pueden incluirse por medio de la identificación de cada cuadro con un contador binario (palabra). En la figura 48-55 se ilustra un formato típico de cuadro de datos para telemetría (TDM). La frecuencia de muestreo y longitud de palabra del sistema PCM determinan la velocidad de transmisión de datos TDM que debe regir en el canal.

Multicanalización por portadora en cuadratura. Aun cuando no se utiliza tanto como los sistemas FDM y TDM, una señal portadora en cuadratura puede emplearse para la transmisión simultánea de dos canales de información. La señalización por cuadratura se ha analizado ampliamente en términos de la denominada modulación de fase en cuadratura o modulación por cuadratura de fase, en la que el canal de información es una sola corriente de datos digitales que se reduce en frecuencia en un factor de dos, con la mitad de la información modulada en la componente en fase y la otra mitad en la componente en cuadratura de fase de la portadora. Los datos desmodulados de cada componente de la portadora se vuelven a combinar a fin de obtener la corriente de datos original. La ventaja evidente en este caso es la reducida velocidad de transmisión de datos, lo que permite que los datos de mayor velocidad sean transferidos por un canal de ancho de banda limitado. Sin embargo, no es necesario que ambas entradas de datos se sinteticen una de otra en cual-

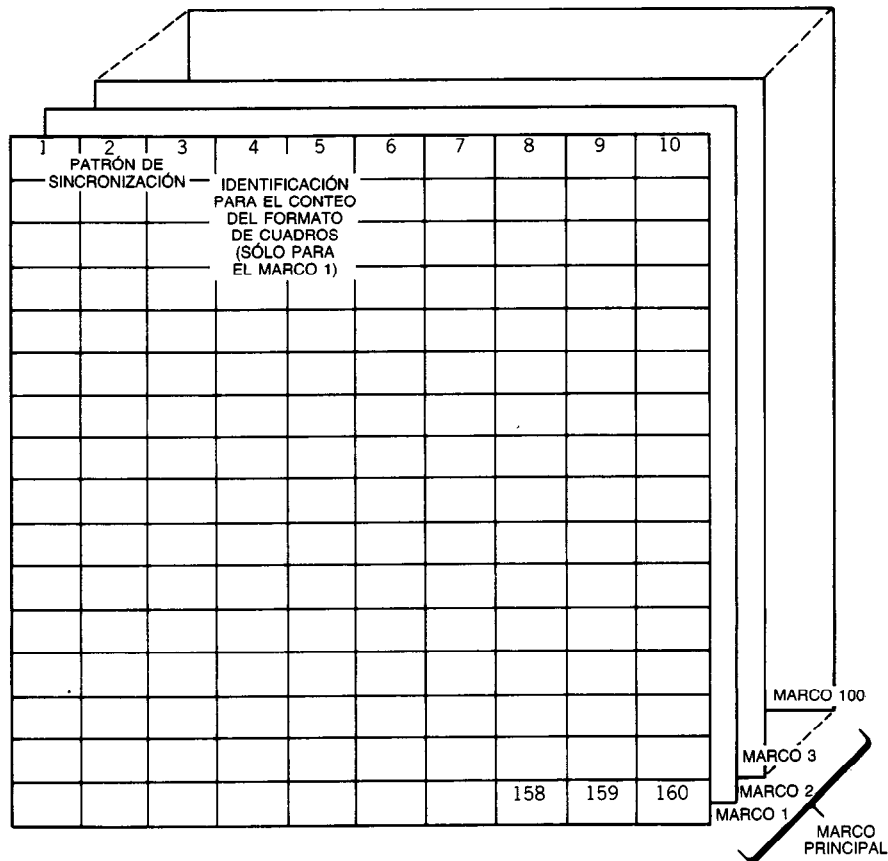


Fig. 48-55. Formato típico de cuadros de datos para telemetría por modulación por codificación de pulsos (PCM), para datos de multicanalización por división en el tiempo (TDM). Patrón de sincronización: 76571440₈. 128 kilobits por segundo, 8 bits por palabra, 160 palabras por cuadro secundario, 100 cuadros secundarios por cuadro principal, y 1 cuadro principal por segundo.

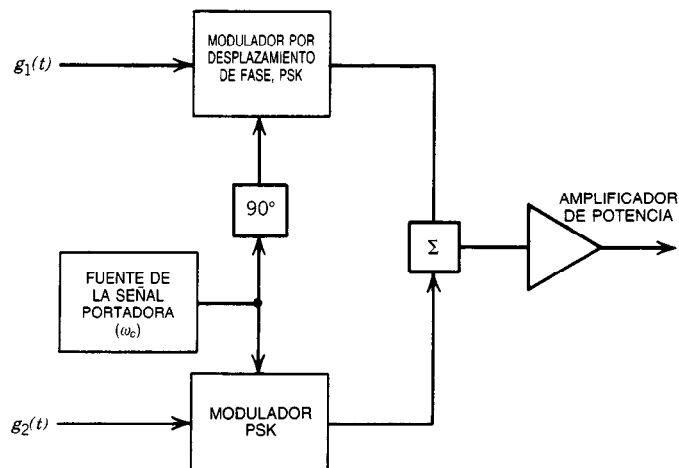


Fig. 48-56. Multicanalización por portadora en cuadratura, ilustración funcional.

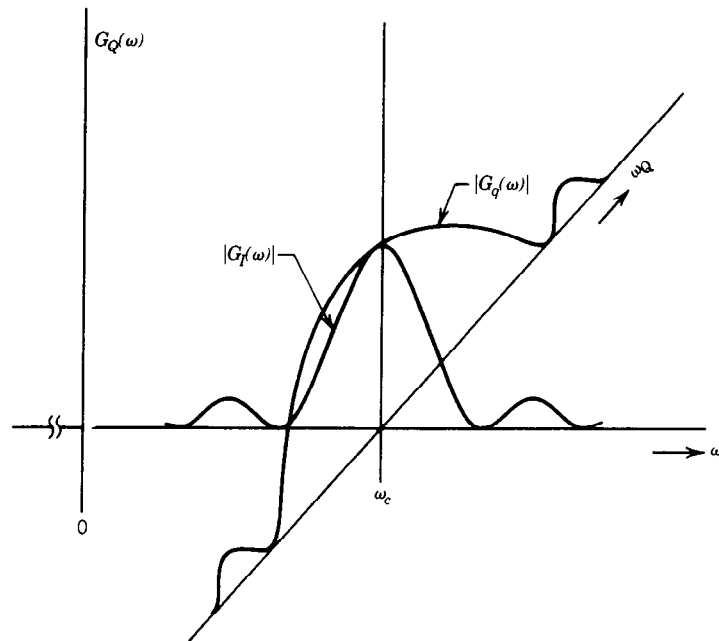


Fig. 48-57. Espectros de frecuencia de la multicanalización por portadora en cuadratura.

quier sentido. En este caso, las señales en cuadratura y en fase podrían servir como portadoras por separado de la misma frecuencia para la transmisión simultánea de dos canales de información. En la figura 48-56 se muestra una mecanización conceptual de la multicanalización por portadora en cuadratura, y en la figura 48-57 un espectro de potencia de modulación típico. Los espectros $G_f(\omega)$ y $G_q(\omega)$ son los espectros resultantes de la modulación de datos para las señales en fase y en cuadratura, respectivamente (Lindsey).

Una forma de multicanalización más complicada que ha surgido recientemente es la denominada multicanalización por división de códigos. En este método las señales de información normalmente se transmiten de forma simultánea desde fuentes separadas de transmisión, y son recibidas por un solo elemento receptor. En la figura 48-58 se ilustra el sistema. Cada transmisor de señales de información es modulado con una sucesión pseudoaleatoria de códigos (PRN, de *pseudorandom*) antes de la transmisión. En el receptor, las señales codificadas combinadas se convierten en una señal IF cuyo ancho de banda sea lo suficientemente amplio para pasar los espectros codificados. Luego, cada uno de los códigos se detecta y sincroniza con la réplica del receptor de la sucesión PRN transmitida. Esta réplica de código se utiliza para desmodular la señal IF que produce la versión recibida de dicha señal de información del código en la IF. Las otras señales codificadas permanecen como ruido de fondo o interferencia adicional, con lo que se establece un límite superior con respecto al número total de señales codificadas que es posible recibir de forma simultánea. Después se desmodula el canal codificado a fin de obtener la señal de información $g(t)$.

La multicanalización por división de códigos es un ejemplo de las técnicas de espectro extendido, que representan un avance importante en el área de las comunicaciones digitales. Estas técnicas se describen con más detalle en la sección 48-5.

LIBROS Y REVISTAS

- B. H. Batson, *An Analysis of the Relative Merits of Various PCM Code Formats*, JSC Internal Note MSC-EB-R-68-5, pp. 3-1-4-11.
- B. H. Batson, "Performance of Binary FSK Data Transmission Systems", Informe de Johnson Space Center, JSC 08097.
- J. J. Downing, *Modulation System and Noise*, Prentice-Hall, Englewood Cliffs, NJ, pp. 113-177.
- L. J. Giacoletto, "Generalized Theory of Multitone Amplitude and Frequency Modulation", *Proc. IRE*, pp. 680-693.
- P. M. Hopkins y M. H. Kapell, "A Correction Factor for Filter Degradation in Frequency Modulation System Performance Models", Informe de Johnson Space Center, JSC-09059.
- W. C. Lindsey, *Synchronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, NJ.
- W. C. Lindsey, "Power Division Analysis of a Three Channel Unbalanced QPSK Signal Out of a Band Pass Filter", LinCom Corp. TR n.º 04-7604-8.
- J. H. Painter y G. Hondros, "Unified S-Band Telecommunication Techniques for Apollo", Vol. II, *Mathematical Models and Analysis*, NASA, # 1184.

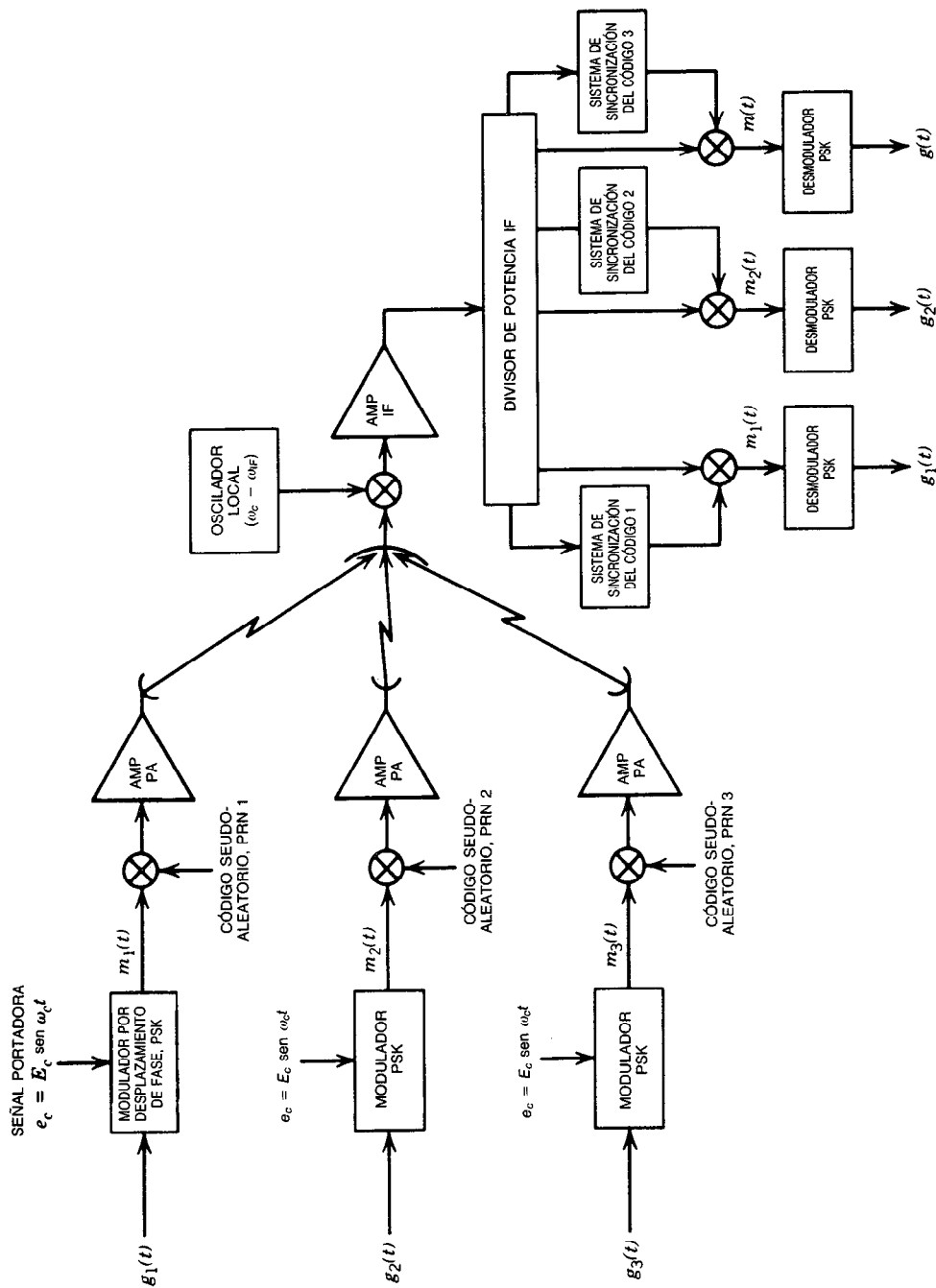


Fig. 48-58. Sistema de multicanalización por división de códigos, diagrama funcional.

- R. J. Panneton, *Analysis of Reduction in the 1.024 MHz Telemetry Subcarrier Modulation Index*, TRW, Informe de Proyecto Técnico Task E-94.
- P. F. Panter, *Modulation, Noise, and Spectral Analysis*, McGraw-Hill, Nueva York.
- S. C. Platkin, "FM Bandwidth as a Function of Distribution and Modulation Index", *IEEE Communications Technology*.
- M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, Nueva York.
- J. W. Seyl, B. H. Batson y B. G. Smith, "Experimental Result for FSK Data Transmission Systems Using Discriminator Detection", Proc., National Telecommunication Conference.
- J. W. Seyl y colaboradores, "Shuttle S-Band Communications Technical Concepts", Proc., Space Shuttle Technical Conference, Johnson Space Center.
- D. Shilling y M. Taub, *Principles of Communications Systems*, McGraw-Hill, Nueva York.
- R. C. Tittsworth y L. R. Welch, "Power Spectra of Signals Modulated by Random and Pseudo Random Sequence", Informe Técnico 32-140, Jet Propulsion Laboratory, Pasadena, CA.
- A. J. Viterbi, *Principles of Coherent Communications*, McGraw-Hill, Nueva York, pp. 185-215.

48.2 RADIOCOMUNICACIONES

George W. Raffoul

48.2.1 Introducción

La radiocomunicación es un proceso por medio del cual es posible transmitir cualquier tipo de información de un sitio a otro distante sin tener que tender cables entre la fuente y el destino. La transmisión de formas de comunicación, tales como las señales de audio, vídeo o telemetría, se logra modulando la amplitud, frecuencia o fase (o alguna combinación de ellas) de una portadora de frecuencia superior. La portadora modulada se transforma en una onda electromagnética de propagación, que se desplaza por un medio de propagación (por lo general la atmósfera) a fin de llegar al destino de la información. En el destino la onda se intercepta y se transforma de vuelta en una señal portadora modulada, se amplifica y se desmodula a fin de recuperar la información original.

Por tanto, el proceso de radiocomunicación implica una serie de pasos consecutivos para el transporte de la información desde la fuente al destino.

- **Paso uno.** Consiste en la modulación de la portadora en la fuente por la señal de información. Esta función es realizada por el transmisor.
- **Paso dos.** Implica la traducción de la portadora modulada en la onda electromagnética de propagación por la antena transmisora.
- **Paso tres.** Consiste en la propagación de la onda electromagnética a través de la atmósfera hasta que llega al destino deseado.

- **Paso cuatro.** La antena receptora convierte la onda electromagnética recibida en el destino en una señal portadora modulada.
- **Paso cinco.** Consiste en la desmodulación o extracción de la información contenida en la portadora modulada una vez que se ha amplificado lo suficiente.

La amplificación es realizada por el receptor, cuya capacidad para recuperar la información original con exactitud es estorbada por la adición de ruido en el canal. El rendimiento global del enlace de radiocomunicación depende no sólo del contenido de ruido, sino también del tipo de modulación y las técnicas de codificación utilizados. En un sistema de comunicación analógico, el rendimiento se expresa en términos de la razón de señal sobre ruido (SNR, de *signal-to-noise ratio*) posdetección en el receptor. Sin embargo, en un sistema de comunicación digital, en el cual el objetivo es decidir si se ha transmitido un 1 o un 0, el rendimiento suele describirse en términos de la probabilidad de error en los bits, P_B , a la salida del detector.

En esta sección se presenta un panorama general de cada uno de los cuatro componentes de un enlace de radiocomunicación, junto con un breve análisis del rendimiento del enlace. En el tema de la propagación de ondas de radio, se destacan especialmente los efectos atmosféricos. La tendencia actual hacia las frecuencias de ondas milimétricas hace que el impacto de la absorción por vapor de agua y oxígeno sea lo suficientemente importante para ser incluida en el cálculo del margen del enlace.

En el dominio de la antena, los esfuerzos actuales de desarrollo están orientados hacia las redes de antenas en fase barridas electrónicamente, que pueden incluir características avanzadas, tales como el control adaptativo y capacidad de conformación a las superficies curvas. También se está dedicando bastante esfuerzo de desarrollo a las antenas de haces múltiples de satélite, las antenas de microcinta de bajo perfil y las antenas para ondas milimétricas.

Con respecto al tema de los transmisores y receptores, en esta sección, el análisis está limitado casi exclusivamente a los sistemas analógicos clásicos, como los de AM (modulación de amplitud), FM (modulación de frecuencia) y SSB (de banda lateral única). Los sistemas de comunicación digital son en la actualidad los que mayor atención están recibiendo en el campo de la investigación y en el mercado. En la sección 48.5 se cubren con mayor detalle las técnicas de comunicación digital.

El último tema que se aborda en esta sección está relacionado con el análisis de enlaces de radiocomunicación. Se proporciona una fórmula para el margen de enlace, y su aplicación en el cálculo del presupuesto de enlace se ilustra con un ejemplo.

48.2.2 Propagación de ondas de radio

Propagación en el espacio libre

El **espacio libre** se define idealmente como un medio homogéneo sin corrientes o cargas conductoras pre-

Tabla 48-7. Espectro electromagnético

Designación de la banda	Intervalo de frecuencia	Intervalo de longitud de onda en el espacio libre
Extremadamente baja frecuencia (ELF)	< 3 kHz	> 100 km
Muy baja frecuencia (VLF)	3 – 30 kHz	10 – 100 km
Baja frecuencia (LF)	30 – 300 kHz	1 – 10 km
Frecuencia media (MF)	300 kHz – 3 MHz	100 m – 1 km
Alta frecuencia (HF)	3 – 30 MHz	10 – 100 m
Muy alta frecuencia (VHF)	30 – 300 MHz	1 – 10 m
Ultraalta frecuencia (UHF)	300 MHz – 3 GHz	10 cm – 1 m
Superalta frecuencia (SHF)	3 – 30 GHz	1 – 10 cm
Extremadamente alta frecuencia (EHF)	30 – 300 GHz	1 – 10 mm
Infrarrojo	$8 \times 10^{11} - 4 \times 10^{14}$ Hz	80 – 400 μ m
Luz visible	$4 \times 10^{14} - 7.5 \times 10^{14}$ Hz	40 – 80 μ m
Luz ultravioleta	$7.5 \times 10^{14} - 10^{16}$ Hz	1.2 – 40 μ m
Rayos X, rayos gamma	$10^{16} - 10^{20}$ Hz	0.6 μ m – 1.2 μ m
Rayos cósmicos	$> 10^{20}$ Hz	< 0.6 m μ

sentos, y sin objetos que absorban o reflejen energía radioeléctrica. Este concepto se utiliza debido a que simplifica el entendimiento de la propagación de ondas y porque las condiciones de propagación algunas veces se aproximan a las del espacio libre, especialmente a frecuencias en la región superior de ultraalta frecuencia. El alcance de propagación de las ondas de radio en el espacio libre solamente es limitado por la atenuación de la señal a medida que la onda se desplaza alejándose de la fuente de radiación. La pérdida de transmisión o de trayectoria está dada por la ley del cuadrado inverso de la óptica aplicada a la radiotransmisión. La atenuación de la trayectoria entre dos antenas isotrópicas está dada por:

$$L_p = \left(\frac{4\pi d}{\lambda} \right)^2 \quad (48-80)$$

donde L_p = pérdida de trayectoria, o razón numérica de la potencia transmitida entre la potencia recibida (P_t/P_r)

λ = longitud de onda

d = longitud de la trayectoria

y donde las cantidades físicas tienen las mismas unidades. Una forma más conveniente para la expresión de la pérdida de trayectoria es:

$$L_p = 32.5 + 20 \log f + 20 \log d \quad (48-81)$$

donde L_p = pérdida de trayectoria, dB

f = frecuencia, MHz

d = longitud de la trayectoria, km.

Espectro electromagnético

La velocidad de propagación de las ondas electromagnéticas en el espacio libre es de importancia fundamental. Este valor es igual a la velocidad de la luz en el espacio libre, designada por la letra c . El valor de c es

186 283 millas terrestres por segundo, o 299 793 km/s, que para la generalidad de los casos se redondea a 186 000 mi/s o a 3×10^8 m/s. En medios diferentes al espacio libre, la velocidad de las ondas electromagnéticas suele ser menor, dependiendo de las características del medio de propagación en cuestión. Otro parámetro de las ondas electromagnéticas es la longitud de onda, que es igual a la distancia que recorre la onda durante el periodo de un ciclo. Se denota λ_0 para el espacio libre y está relacionada con la velocidad por la expresión:

$$\lambda_0 = \frac{c}{f} \quad (48-82)$$

donde la longitud de onda λ_0 está en metros, la velocidad c en metros por segundo y la frecuencia f en ciclos por segundo, o hertz. Ya que los diversos fenómenos de propagación de ondas dependen en gran parte de la frecuencia de la onda, en la tabla 48-7 se muestra todo el espectro electromagnético. Obsérvese la subdivisión del espectro de radio en varias bandas, conforme a la costumbre aceptada. Las frecuencias y longitudes de onda se presentan en las unidades de radio usuales.

Efectos atmosféricos

En el concepto de transmisión en el espacio libre se supone que la atmósfera es perfectamente uniforme y no absorbente, y que la Tierra se encuentra infinitamente lejos o que su coeficiente de reflexión es despreciable. En la práctica, durante la propagación cerca de la Tierra, las ondas se reflejan en suelo, montañas y construcciones; son refractadas a medida que pasan a través de capas atmosféricas de diferentes densidades o diferentes grados de ionización. Asimismo, las ondas electromagnéticas pueden difractarse alrededor de grandes obstáculos e incluso pueden interferir entre sí, como cuando dos ondas provenientes de la misma

fuentes se encuentran después de haber recorrido trayectorias diferentes. Las ondas también pueden ser absorbidas por diversos átomos y moléculas que se encuentren en la atmósfera. Algunos de estos efectos son deseables, hasta cierto punto, en el caso de enlaces de muy baja frecuencia (VLF, de *very low frequency*), baja frecuencia (LF, de *low frequency*), mediana frecuencia (MF, de *medium frequency*), alta frecuencia (HF, de *high frequency*) y por tropodispersión, debido a su ayuda en el establecimiento de una trayectoria de comunicación entre el transmisor y el receptor. Sin embargo, en un enlace de comunicación en línea recta, esos efectos dificultan la función de comunicación y dan por resultado desvanecimiento temporal o simplemente atenuación en el nivel de la portadora. A continuación se analizará brevemente cada uno de estos efectos atmosféricos.

Absorción atmosférica. De los principales gases presentes en la atmósfera, el vapor de agua y el oxígeno son los que absorben más energía de las ondas electromagnéticas; el primero debido a su momento dipolar eléctrico y el segundo debido a su momento dipolar magnético. Por fortuna, la absorción atmosférica a frecuencias por debajo de 10 GHz es bastante insignifi-

cante. Como se muestra en la figura 48-59, la absorción tanto por el contenido de oxígeno como por el de vapor de agua en la atmósfera se vuelve importante a esa frecuencia y después aumenta de manera gradual. Sin embargo, debido a las diferentes resonancias moleculares existen ciertas crestas y valles de atenuación. Por ejemplo, para la comunicación atmosférica a larga distancia no se recomiendan frecuencias como 60 y 120 GHz. Asimismo, es mejor no utilizar 23 o 180 GHz, excepto en ambientes de aire fresco. Por otra parte, en la figura 48-59 se muestran ventanas o bandas de frecuencia donde la absorción se reduce bastante, como en el caso de frecuencias en la vecindad de 33 y 110 GHz. Obsérvese además que el nivel de absorción debido al contenido de vapor de agua está basado en un nivel de humedad medio o estándar. Si éste se incrementa, o si existe niebla, lluvia o nieve, esta forma de absorción aumenta enormemente, y también es posible que se presente dispersión.

Refracción atmosférica. Los fenómenos de refracción atmosférica que se observan en la radiopropagación se deben a variaciones en el índice de refracción del aire a medida que aumenta la altitud. En condiciones normales, el índice de refracción de la atmósfera decrece

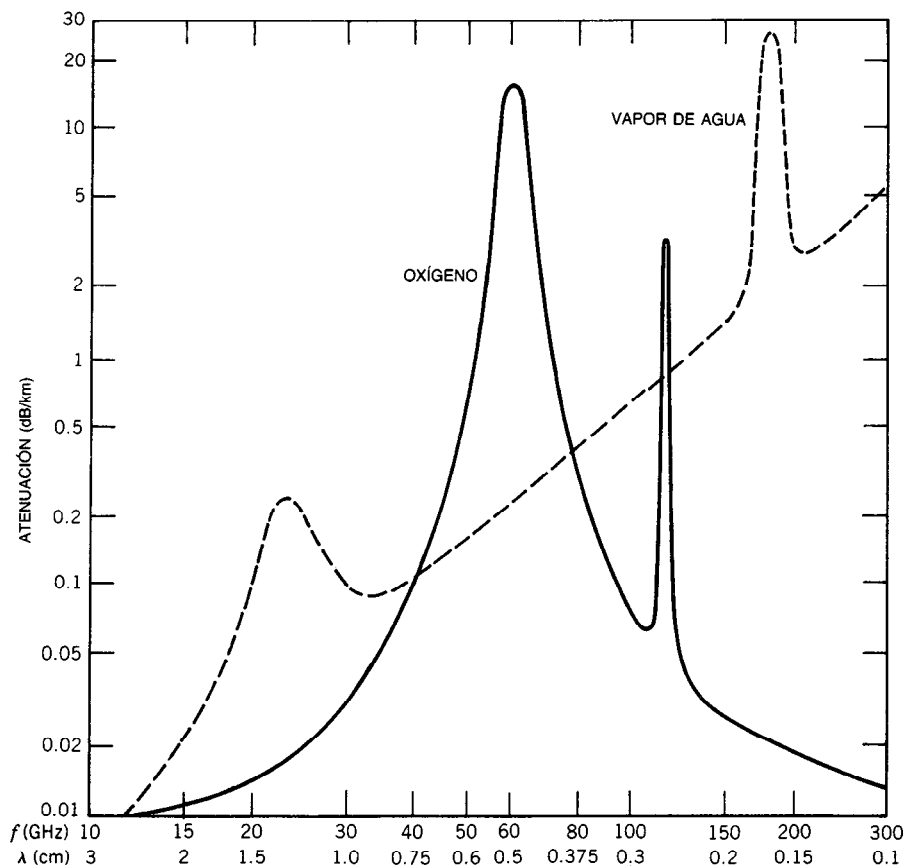


Fig. 48-59. Absorción atmosférica de ondas electromagnéticas. Tomado de Kennedy¹, con permiso.

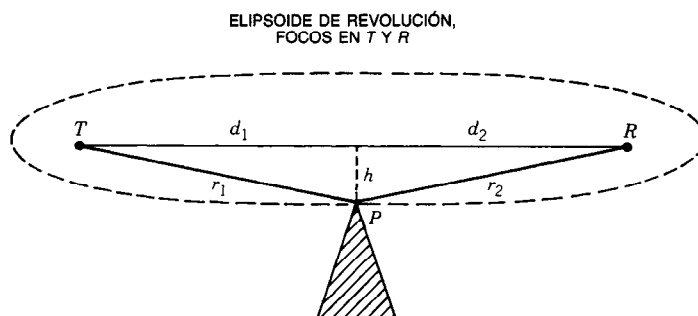


Fig. 48-60. Efecto de difracción de una gran cumbre. T = transmisor; R = receptor.

ligeramente, aunque de forma lineal, con la altitud. Como resultado de la ligera refracción que ocurre, las ondas se curvan un poco en vez de desplazarse estrictamente en líneas rectas, y el radiohorizonte u horizonte radioeléctrico aumenta debido a ello. Básicamente, lo que ocurre es que la parte inferior del frente de onda se desplaza en una atmósfera más densa que la parte superior, por lo que se mueve con más lentitud, debido a lo cual el frente de onda se curva hacia abajo.

La flexión uniforme puede representarse mediante la propagación en línea recta, pero con el radio de la Tierra modificado, de modo que la curvatura relativa entre los rayos radiales y la Tierra no se modifique. El nuevo radio de la Tierra se denomina radio terrestre efectivo, y la razón entre los radios efectivo y real de la Tierra se designa por K . El valor medio de K es aproximadamente $4/3$. La distancia al radiohorizonte sobre la Tierra lisa, cuando la altitud h es muy pequeña en comparación con el radio terrestre, está dada por:

$$d_h = \left(\frac{3Kh}{2} \right)^{1/2} \quad (48-83)$$

donde d_h es la distancia al radiohorizonte en millas y h es la altitud por encima de la superficie terrestre, en pies².

Difracción por obstáculos. Las señales que se propagan sobre una línea recta pueden recibirse en la parte posterior de grandes edificios, montañas y otros obstáculos semejantes, como resultado de la difracción. El fenómeno de la difracción se explica mejor aplicando los conceptos introducidos por Fresnel en el terreno de la óptica. Considérese una gran cumbre localizada a distancias d_1 y d_2 del transmisor T y del receptor R , respectivamente, a una altura por debajo o por encima de la línea recta óptica, como se muestra en la figura 48-60. Las regiones espaciales correspondientes a longitudes de trayectorias incrementadas de múltiplos de $\lambda/2$ con respecto a $TR = d_1 + d_2 = d$ son elipsoides cuyos efectos se encuentran en T y R . La n -ésima superficie de Fresnel es aquella para la cual la suma de las distancias entre el transmisor y el receptor y un punto sobre la superficie del elipsoide de revolución (línea discontinua en la fig. 48-60) exceda por $n(\lambda/2)$ la distancia entre el transmisor y el receptor:

$$S = (r_1 + r_2) - (d_1 + d_2) = \frac{n\lambda}{2} \quad (48-84)$$

Cuando h es pequeña con respecto a d_1 y d_2 , la distancia S puede expresarse como:

$$S = \frac{h^2}{2} \left(\frac{1}{d_1} + \frac{1}{d_2} \right) = \frac{n\lambda}{2} \quad (48-85)$$

Para cada valor entero n existe un valor h_n , denominado radio de la n -ésima región de Fresnel, que está dado por:

$$h_n = \left(\frac{n\lambda}{1/d_1 + 1/d_2} \right)^{1/2} \quad (48-86)$$

Un criterio para determinar si la cumbre está lo suficientemente alejada de la línea recta radioeléctrica entre el transmisor y el receptor, para permitir aplicar las condiciones medias de propagación en el espacio libre, es que la primera región de Fresnel se encuentre libre de obstáculos en la trayectoria de los rayos. El espacio libre de la primera región de Fresnel está dado por:

$$h_1 = \left(\frac{\lambda}{1/d_1 + 1/d_2} \right)^{1/2} \quad (48-87)$$

En la figura 48-61 se muestra el efecto del espacio libre de la trayectoria en la radiotransmisión.

Reflexiones por la superficie terrestre. Las reflexiones debidas a la superficie terrestre pueden afectar seriamente el nivel de la señal recibida.

El caso más sencillo por considerar es la reflexión debida a una superficie terrestre plana, como se muestra en la figura 48-62. La onda resultante en el receptor R consta del rayo dirigido TR y del rayo reflejado OR . La reflexión provoca atenuación y defasamiento a lo largo del rayo indirecto (reflejado), lo cual puede explicarse por medio del coeficiente de reflexión (que expresa la relación de la intensidad de campo del rayo reflejado entre la intensidad de campo del rayo incidente):

$$\Gamma = \rho e^{j\psi} \quad (48-88)$$

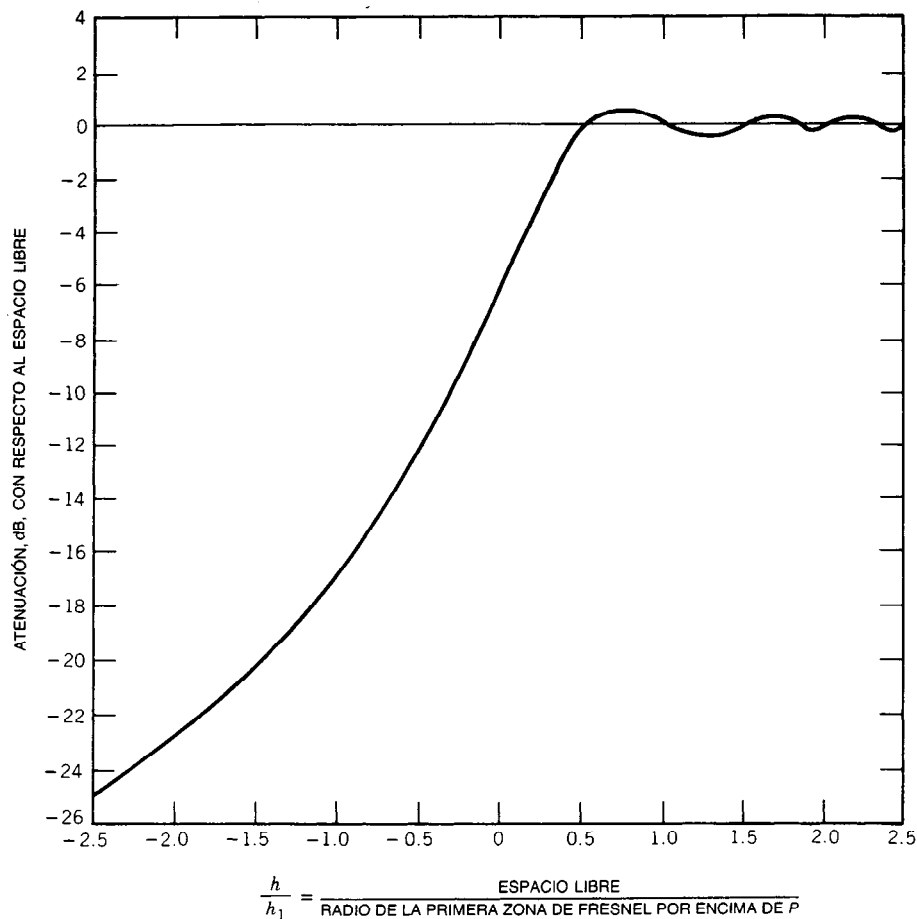


Fig. 48-61. Efecto del espacio libre de la trayectoria en la radiotransmisión, suponiendo una difracción de cuña. Tomado de Panter³, con permiso.

donde ρ y ψ son la magnitud y fase del coeficiente de reflexión, respectivamente. La diferencia de trayectoria entre los rayos reflejado (indirecto) y directo será:

$$S = (TO + OR) - TR \quad (48-89)$$

En la práctica, h_t y h_r son pequeños en comparación con d , así que la diferencia de trayectoria puede aproximarse por:

$$S \approx \frac{2h_t h_r}{d} \quad (48-90)$$

Por consiguiente, el receptor recibe, además de la onda directa, una onda reflejada con amplitud relativa ρ y defasamiento relativo α , que es el resultado del retraso debido a la diferencia de trayectoria S , junto con el defasamiento ψ debido a la reflexión. El defasamiento total está dado por:

$$\alpha = \psi - \frac{2\pi S}{\lambda} = \psi - \frac{4\pi h_t h_r}{\lambda d} \quad (48-91)$$

Para ondas polarizadas horizontales, ρ es casi la unidad y ψ es igual a 180° para casi todos los tipos de terreno y ángulos de incidencia. Para ondas polarizadas verticalmente, ρ y ψ varían considerablemente con las propiedades del terreno y el ángulo de incidencia. En cualquier caso, la amplitud de la onda resultante en la entrada del receptor varía, ya que h_t o h_r varían, entre $1 - \rho$ y $1 + \rho$. Si ρ es casi la unidad, la onda directa puede ser cancelada casi por completo por la onda reflejada (indirecta). Las variaciones en intensidad de la señal debidas a transmisión en múltiples trayectorias pueden reducirse mediante la recepción ya sea en diversidad de frecuencias o en diversidad de espacio.

Tipos de ondas de propagación en la atmósfera

En general, las ondas se desplazan en línea recta, excepto donde la presencia de la Tierra y su atmósfera tienden a modificar la trayectoria. Por consiguiente, si el efecto de la refracción provocada por densidad atmosférica cambiante no se toma en cuenta, es posible afirmar que las frecuencias por encima del intervalo de

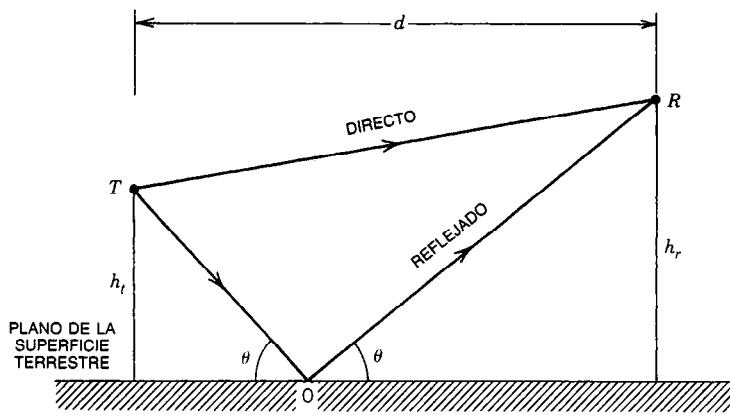


Fig. 48-62. Reflexión debida a una superficie terrestre plana. T = transmisor;
R = receptor.

HF suelen desplazarse en línea recta y propagarse por medio de las denominadas ondas espaciales. Las frecuencias por debajo de este intervalo se desplazan siguiendo la curvatura de la Tierra mediante un efecto de guía de onda acoplado con difracción. La superficie de la Tierra y la capa ionizada más baja de la atmósfera forman los dos muros de la guía de onda. La onda de propagación se denomina, en este caso, onda terrestre o superficial.

Las ondas en el intervalo HF son reflejadas por las capas ionizadas de la atmósfera y se denominan ondas celestes o espaciales. Las señales en la banda HF se transmiten hacia el cielo, donde son reflejadas por las capas ionosféricas y regresan a la Tierra mucho más allá del horizonte. Una comunicación a larga distancia, popularmente conocida como onda corta, es posible cuando estas ondas son reflejadas de forma alternada una y otra vez por la Tierra y la ionosfera hasta llegar a receptores situados en lados opuestos del planeta.

Otros dos métodos de propagación más allá del horizonte son la dispersión troposférica o tropodispersión

y la comunicación geosíncrona vía satélite. El carácter especial de la propagación HF merece tratamiento diferenciado, por lo que se describirá a continuación con mayor detalle.

Propagación HF. La ionosfera es la región muy poco densa que se extiende más o menos desde 50 km hasta 400 km de altitud. Durante el día está bombardeada por la radiación solar, de modo que el aire rarificado se ioniza por efecto de las emisiones de rayos ultravioleta y rayos X. Diferentes partes del espectro solar son absorbidas a distintas altitudes, de modo que se forman varias capas ionizadas. Las capas superiores reflejan ondas de radio, mientras que las inferiores atenúan las ondas que pasan a través de ellas, como se ilustra en la figura 48-63.

El índice de refracción radioeléctrico μ para una capa ionizada está dado de forma aproximada por:

$$\mu = \left(1 - \frac{f_p^2}{f^2}\right)^{1/2} \quad (48-92)$$

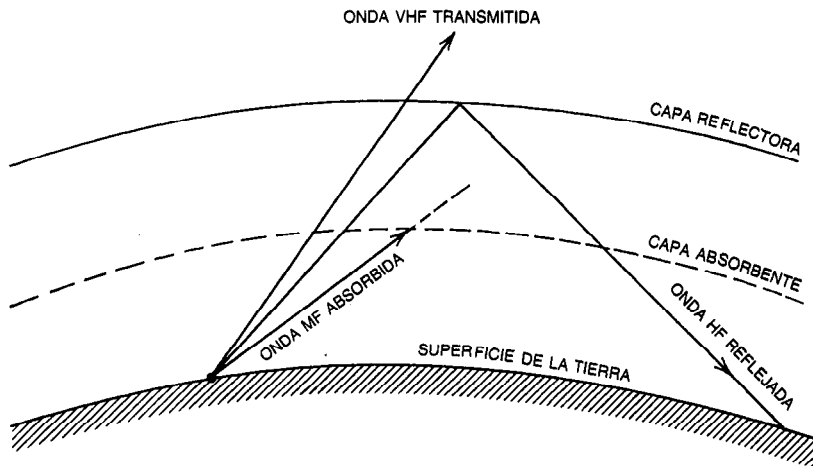


Fig. 48-63. Reflexión de ondas de alta frecuencia por la ionosfera.

donde f_p es la frecuencia del plasma y f es la frecuencia de la onda. Para cualquier densidad dada de ionización N (electrones por metro cúbico), el índice de refracción decrece a medida que disminuye la frecuencia f . Para $f = f_p = 81 N$, el índice de refracción tiende a cero. Una onda de radio a una frecuencia por debajo de f_p , que incide normalmente en una capa ionizada, penetra sólo hasta el punto en que μ se vuelve cero, y entonces es reflejada por completo. La frecuencia del plasma en la parte más elevada de la capa se denomina frecuencia crítica, f_c . Las ondas que inciden normalmente a frecuencias por encima de f_c atraviesan toda la capa.

A frecuencias por encima de la frecuencia crítica, la reflexión aún puede efectuarse si la onda incide oblicuamente. La máxima frecuencia que es posible reflejar por completo a un ángulo dado de incidencia θ se denomina frecuencia máxima utilizable, o MUF (de *maximum usable frequency*), y está dada por:

$$\text{MUF} = f_c \sec \theta \quad (48-93)$$

La MUF puede ser varias veces mayor que la frecuencia crítica, y su expresión es solamente aproximada debido a que se basa en una superficie terrestre y una capa reflectora planas.

La pérdida por atenuación depende principalmente del producto de la densidad de ionización y la frecuencia de choques entre los electrones libres y partículas más pesadas. Cuando los electrones son puestos en movimiento por una onda de radio, adquieren energía oscilatoria de la onda, que se convierte en calor durante los choques. Dado que la energía se obtiene de la onda, ésta es atenuada. El grado (nivel) de atenuación A está dado aproximadamente por:

$$A = \frac{1.16 \times 10^{-6} N \nu}{f^2} \quad \text{dB/m} \quad (48-94)$$

donde ν es la frecuencia de choques de los electrones. Por consiguiente, el nivel de atenuación determina la frecuencia mínima utilizable, mientras que la MUF establece el límite superior.

Dado que la densidad de ionización de la atmósfera superior puede calcularse aproximadamente a partir de la hora, estación, latitud y número de manchas solares, es posible predecir frecuencias y atenuaciones críticas para la mayor parte de los fines prácticos, excepto durante el pico del ciclo solar de 11 años. En este periodo crítico la actividad solar es muy intensa, y podría producirse una pérdida total de la comunicación HF en toda la banda. En la figura 48-64 se presentan las altitudes de las diversas capas ionosféricas con sus variaciones regulares. La capa D es estrictamente una capa de atenuación. Las capas E , F_1 y F_2 son capas reflectoras durante el día. Por la noche, las capas D y E desaparecen y las capas F_1 y F_2 se combinan en la capa F , a la altitud de la capa F_1 .

48.2.3 Fundamentos sobre antenas

Tipos de antenas

Las antenas pueden clasificarse, en términos generales, ya sea por el espectro de frecuencia en que suelen aplicarse o por su modo básico de radiación. En el primer tipo de clasificación, los tipos de antenas obedecen a las designaciones usuales de las bandas y se describen como antenas de VLF, LF, MF, HF, UHF y microondas, conforme a su frecuencia de operación. En el segundo tipo de clasificación, las antenas pueden dividirse en cuatro grupos:

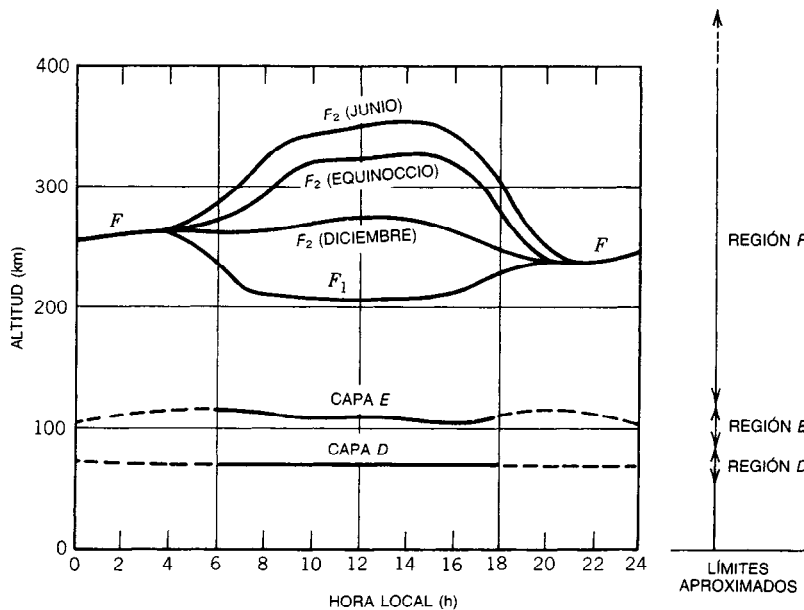


Fig. 48-64. Capas ionosféricas y sus variaciones regulares.

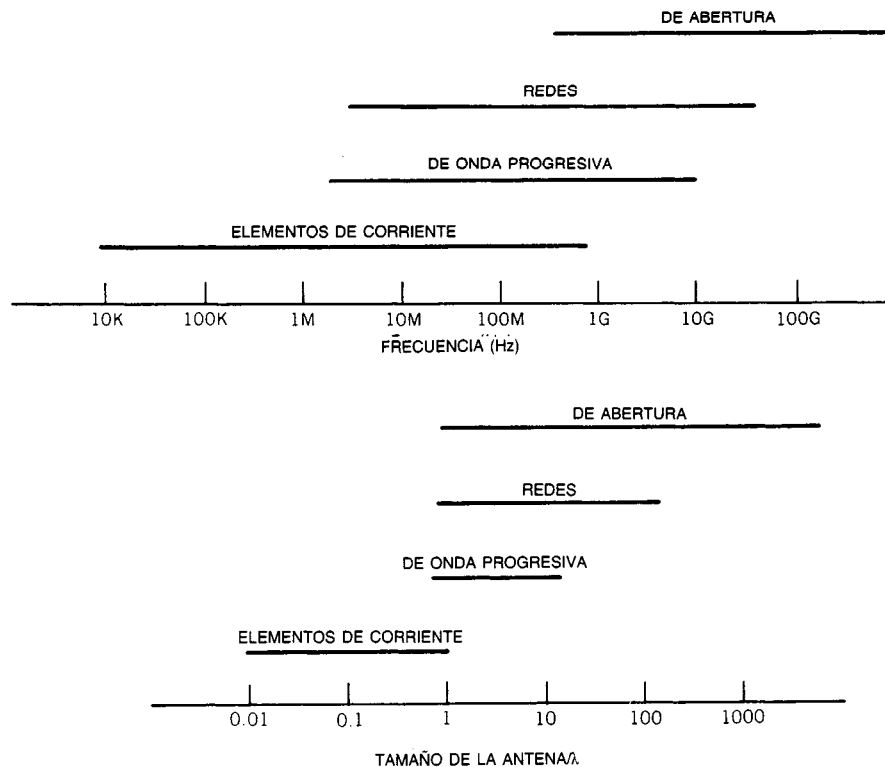


Fig. 48-65. Clasificación de las antenas con base en el modo de radiación. Tomado de Rudge y colaboradores⁴, con permiso.

- De elemento de corriente.
- De onda progresiva.
- Múltiples.
- De abertura.

Los cuatro grupos pueden distinguirse por el tamaño de la antena medido en longitudes de onda, que a su vez puede relacionarse con las distintas regiones del espectro en las cuales suelen aplicarse las antenas, como se muestra en la figura 48-65. En la tabla 48-8 se enumeran ejemplos de antenas que pueden ubicarse en cualquiera de los cuatro grupos. La clasificación de las antenas en estos cuatro grupos es solamente una aproximación, con bastantes excepciones. A pesar de ello, constituye una forma conveniente de organizar el tema de los fundamentos de las antenas.

En el análisis matemático de las antenas, la elección correcta del sistema de coordenadas es a menudo un factor importante para simplificar las expresiones de los campos y corrientes electromagnéticos asociados con el sistema de la antena. Dependiendo de la configuración geométrica implicada, es una práctica común utilizar los sistemas de coordenadas cartesianas, polares, cilíndricas o esféricas. En la figura 48-66 se ilustran las componentes del campo eléctrico radiado, en algún punto del espacio, debido a una fuente puntual localizada en el origen, en términos de coordenadas esféricas.

Parámetros de la antena

Características de radiación. La característica (o diagrama) de radiación es el parámetro más importante de una antena, ya que muchos de los demás parámetros suelen deducirse a partir de éste. Debido al principio de reciprocidad, la característica de radiación de una antena de transmisión es equivalente a la característica de recepción de la misma antena cuando se utiliza en el modo de recepción. Por definición, la característica de radiación representa el cambio de la intensidad del campo eléctrico sobre la superficie de una gran esfera de radio r centrada con respecto a la antena radiante. En coordenadas esféricas, es una gráfica de la intensidad del campo eléctrico $E(\theta, \phi)$ como función de las variables direccionales θ y ϕ . En la práctica, esta característica tridimensional se mide y registra en una serie de patrones bidimensionales. Sin embargo, en el caso de las antenas direccionales de un solo haz, puede obtenerse suficiente información acerca de la característica tridimensional a partir de sólo dos patrones planos bidimensionales que incluyen la dirección máxima del haz principal. Estos patrones planos se denominan características de los planos principales de la antena, tales como el plano xy ($\theta = 90^\circ$) y el plano xz ($\phi = 0^\circ$) que se muestran en la figura 48-66. Para una antena polarizada linealmente, las características de planos principales también pueden denominarse carac-

Tabla 48-8. Tipos de antenas

De elemento de corriente	De onda progresiva	Redes	De abertura
Monopolar	De fuente lineal	Dé radiación transversal	De reflector
Dipolar	De alambre largo	De radiación longitudinal	De bocina
De cuadro	Rómbica	Lineal	De lente
De rendija radiante	De guía de ondas ranurada	Planar	De radiación de retroceso
Bicónica	Espiral	Circular	Varilla dieléctrica corta
De ranura	Helicoidal	Conforme	Bocina parabólica
Esferoidal	Periódica logarítmica	Periódica logarítmica	
De reflector parabólico	De ondas lentas	De procesamiento de señales	
De microcinta	De ondas rápidas		
	De ondas dispersas		
	De ondas superficiales		
	De varilla dieléctrica larga		

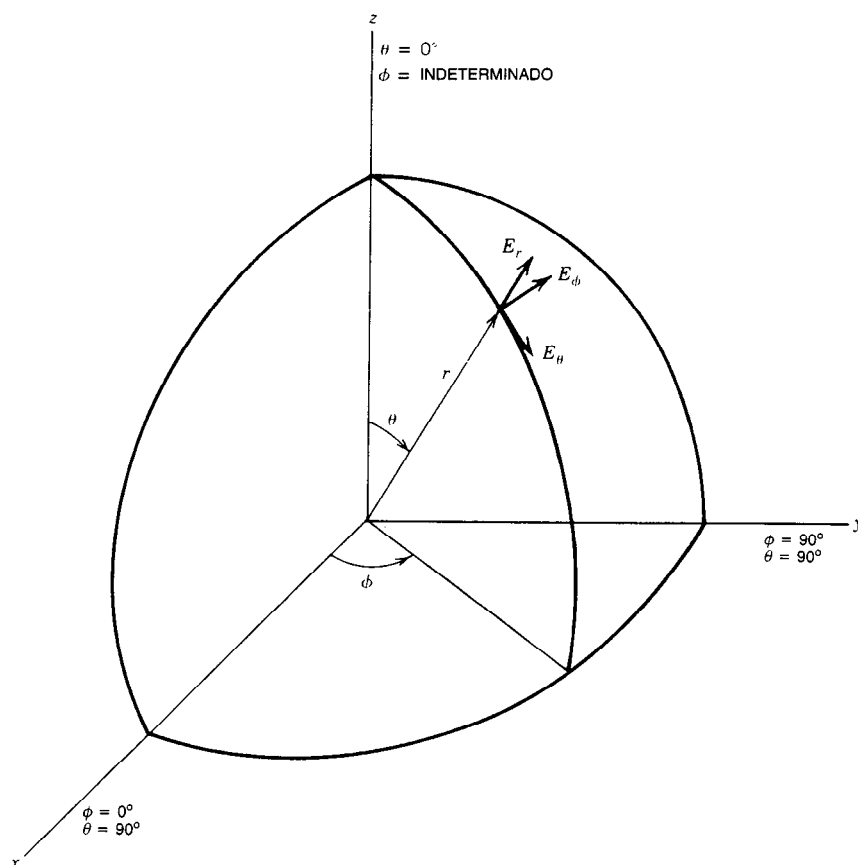


Fig. 48-66. Sistema de coordenadas esféricas utilizado en el análisis de antenas.

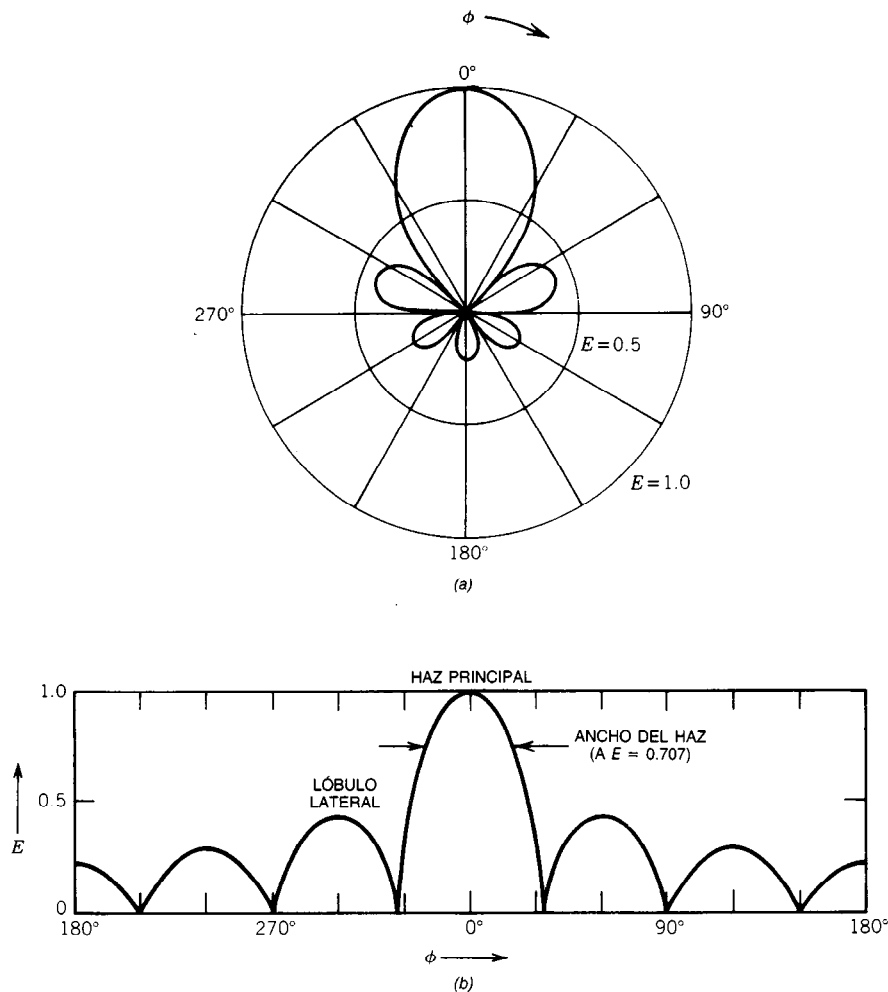


Fig. 48-67. Comparación de diagramas de característica de campo E relativa graficadas en forma polar y rectangular: a) gráfica polar; b) gráfica rectangular.

terísticas de plano E y de plano H , siempre que un plano contenga el vector de campo E y el otro contenga el vector de campo H .

La característica de radiación puede representarse gráficamente en una variedad de formas, de las cuales las más comunes son las gráficas bidimensionales en coordenadas polares o cartesianas. La amplitud relativa de la energía radiada puede registrarse como una característica de potencia relativa ($P/P_{\text{máx}}$), una característica de campo relativa ($E/E_{\text{máx}}$), una característica de potencia logarítmica [$10 \log(P/P_{\text{máx}})$] o una característica de campo logarítmica [$20 \log(E/E_{\text{máx}})$]. En la figura 48-67 se ilustra una característica de campo relativa típica graficada en forma polar y en forma rectangular.

La característica de radiación puede utilizarse para obtener el ancho (o abertura angular) del haz principal y el nivel del lóbulo lateral. El ancho del haz principal se especifica como la diferencia angular entre los

dos puntos en la característica de radiación donde la potencia ha caído a la mitad del valor de cresta, o -3 dB en la escala de decibels. El nivel del lóbulo lateral representa el nivel del mayor lóbulo menor como fracción del nivel del haz principal, y a menudo se especifica en decibels (p. ej., -20 dB).

El espacio que rodea a una antena suele subdividirse en tres regiones o zonas:

1. Región reactiva de campo próximo.
2. Región radiante de campo próximo (de Fresnel).
3. Región de campo lejano (de Fraunhofer).

La primera de estas regiones está muy cerca de la antena, donde las componentes reactivas de los campos electromagnéticos son muy grandes respecto a los campos radiantes. La segunda región se encuentra localizada entre la región reactiva de campo cercano y la

región de campo lejano, donde los campos radiantes predominan y donde la distribución de campo angular depende de la distancia a la antena. La **región de campo lejano** se define como la región donde la distribución de campo angular es esencialmente independiente de la distancia a la antena. Su frontera interior se toma como la distancia radial $R = 2D^2/\lambda$, donde D es la mayor dimensión de la antena y λ es la longitud de onda, mientras que la frontera exterior se localiza en el infinito. Es en esta región donde la característica de radiación de la antena se mide, calcula o ambas cosas.

Directividad, ganancia y eficiencia. La directividad de una antena es una medida de sus propiedades direccionales o de su capacidad de concentrar la potencia radiada en distintas direcciones. Por lo general la directividad se especifica respecto a un radiador isotrópico, que es una antena hipotética que radia uniformemente en todas direcciones. Por tanto, la directividad $D(\theta, \phi)$ en una dirección específica (θ, ϕ) está dada por la razón de la intensidad de la radiación de la antena en la dirección (θ, ϕ) entre la intensidad de la radiación producida por un radiador isotrópico:

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{U_0} = \frac{U(\theta, \phi)}{(P_{\text{rad}}/4\pi)} \quad (48-95)$$

donde $U(\theta, \phi)$ = intensidad de la radiación de la antena en la dirección (θ, ϕ) , W/ángulo sólido unitario

U_0 = intensidad de la radiación de la fuente isotrópica, W/ángulo sólido unitario

P_{rad} = potencia total radiada por la antena, W

En términos de la intensidad del campo eléctrico lejano $E(\theta, \phi)$, la directividad puede expresarse como:

$$D(\theta, \phi) = \frac{|E(\theta, \phi)|^2}{1/4\pi \int_0^{2\pi} \int_0^\pi |E(\theta, \phi)|^2 \sin\theta \, d\theta \, d\phi} \quad (48-96)$$

Aun cuando la directividad puede especificarse en cualquier dirección, se acostumbra hacer referencia al valor de cresta asociado con la dirección del haz principal radiado por la antena. Por consiguiente, en cualquier referencia a la directividad de una antena, el factor de cresta D_0 suele estar implicado. Por ejemplo, un elemento de corriente lineal corto (también denominado dipolo Hertz) tiene directividad de cresta de 1.5 o 1.76 dB con respecto a una fuente isotrópica, mientras que el dipolo de media onda más largo tiene directividad de 1.64 o 2.14 dB.

Para antenas con haz principal estrecho y lóbulos menores bastante despreciables, la directividad de cresta puede aproximarse por medio de:

$$D_0 \approx \frac{41 \, 253}{\theta_{1d} \theta_{2d}} \quad (48-97)$$

donde θ_{1d} es el ancho del haz (en grados) entre los puntos de potencia promedio en un plano principal y θ_{2d} es ese ancho en otro plano principal perpendicular.

Otra medida útil para describir el rendimiento de una antena es la ganancia. Explica no sólo las propiedades direccionales de la antena, sino también su eficiencia. La ganancia de potencia $G(\theta, \phi)$ en una dirección dada se define como la razón de la intensidad de la radiación de la antena en tal dirección, entre la intensidad de la radiación producida por una fuente isotrópica sin pérdidas que tenga la misma potencia total de entrada que la antena:

$$G(\theta, \phi) = \frac{U(\theta, \phi)}{(P_{\text{ent}}/4\pi)} \quad (48-98)$$

donde P_{ent} es la entrada de potencia total aceptada por la antena procedente de un transmisor.

A diferencia de lo que ocurre en el caso de la directividad, en la expresión de la ganancia de potencia se incluye el efecto de las pérdidas óhmicas. Sin embargo, no se incluyen las pérdidas debidas al desacoplamiento de impedancia entre la antena y la línea de transmisión o las pérdidas debidas al desacoplamiento de polarización en una antena receptora. En cualquier caso, la ganancia de potencia es menor que la directividad por un factor igual a la eficiencia de radiación η de la antena:

$$\eta = \frac{G(\theta, \phi)}{D(\theta, \phi)} \quad (48-99)$$

Tanto la ganancia como la directividad pueden estar referidas a cualquier antena normal, tal como un dipolo de media onda, o una bocina en vez de al radiador isotrópico utilizado aquí.

Aun cuando la reciprocidad asegura que los valores calculados de ganancia se aplican igualmente bien a una antena de transmisión o a una antena de recepción, el rendimiento de la segunda también puede describirse en términos de una sección transversal receptora o un área efectiva. Una antena receptora reunirá energía efectiva de una onda plana incidente y, si se acopla adecuadamente, transferirá esta potencia a una carga. La proporción de energía incidente que será destinada a la carga es una función de las propiedades de polarización de la antena y su ganancia en la dirección de la onda plana incidente. La **abertura efectiva de una antena** puede definirse como el área de una antena ideal que podría absorber la misma potencia de una onda plana incidente que la antena en cuestión. El área efectiva de una antena receptora es función del ángulo de llegada de la onda incidente y está relacionada con la ganancia de potencia por:

$$A_{\text{ef}}(\theta, \phi) = \frac{\lambda^2}{4\pi} G(\theta, \phi) \quad (48-100)$$

donde $\lambda^2/4\pi$ es el área efectiva de un radiador isotrópico. La eficiencia de apertura evalúa la apertura efectiva como fracción de la apertura física de la antena. Esta idea es útil para antenas que tienen una apertura de captación bien definida, como es el caso de las antenas de apertura (véanse algunos ejemplos en la tabla 48-8):

$$\eta_{\text{apertura}} = \frac{A_{\text{ef}}}{A_{\text{fis}}} \quad (48-101)$$

Por ejemplo, la eficiencia de apertura de las antenas con reflector se encuentra en el intervalo del 50 al 70%.

Ancho de banda. Este término se utiliza para describir el intervalo de frecuencias sobre el cual una antena funcionará satisfactoriamente. No existe una definición única para el rendimiento satisfactorio, ya que tal rendimiento depende de la aplicación de la antena. Por lo general es posible distinguir entre un ancho de banda determinado por consideraciones de características de radiación y un ancho de banda determinado por consideraciones de impedancia. Asociados con el primer caso están características como la ganancia, nivel del lóbulo lateral, ancho del haz, polarización y dirección del haz, mientras que con el segundo caso se asocian la impedancia de entrada y eficiencia de radiación.

En la práctica, por lo general uno o más parámetros de la antena son más sensibles al cambio de frecuencia que otros, y por tanto pueden constituirse en el factor limitante del ancho de banda. Esto es cierto sólo si la variación de tales parámetros está restringida por los límites de funcionamiento impuestos por la aplicación.

Para antenas de banda amplia el ancho de banda suele expresarse como la razón de las frecuencias superior e inferior en la banda aceptable (p. ej., 10:1). Para antenas de banda estrecha, el ancho de banda se expresa como porcentaje de la frecuencia de la banda central (p. ej., 5%). Los factores físicos del diseño que limitan el ancho de banda varían de una antena a otra. En los elementos monopoles, dipolares, de rana y de microcinta (microbanda), las estructuras son resonantes a frecuencias particulares, y el ancho de banda se determina por las características de impedancia en las terminales de entrada. Por otra parte, los radiadores de bocina (o embudo) están limitados en banda por la naturaleza modal de la propagación de ondas en la estructura de la guía de ondas.

Polarización. La polarización en una onda electromagnética a una sola frecuencia describe la forma del lugar geométrico del extremo del vector de campo eléctrico instantáneo como función del tiempo en una ubicación fija en el espacio, y el sentido en que se traza el lugar geométrico según se observa a lo largo de la dirección de propagación. Un solo elemento de corriente orientado a lo largo del eje x radiará una onda linealmente polarizada con un vector de campo eléctrico orientado en la dirección x . Una antena más complicada puede radiar una onda cuyo vector de campo eléctrico tiene componentes x e y . Si las dos componentes E_x y E_y difieren en fase por 0 o 180°, la onda seguirá estando polarizada linealmente. Si las dos componentes tienen igual magnitud y diferencia de fase de $\pm 90^\circ$, el vector de campo eléctrico resultante en un punto dado del espacio girará a velocidad angular ω , de tal modo que su extremo describirá una circunferencia. En este caso se dice que la onda está polarizada circularmente. En

general, si las dos componentes tienen amplitudes y diferencia de fase arbitrarias, el campo eléctrico instantáneo describirá una elipse y la onda se denominará polarizada elípticamente. Además, cuando la rotación alrededor de la elipse o circunferencia es en sentido horario, la polarización se denomina a la derecha (dextrógira); en caso contrario, se denomina a la izquierda (levógira).

Si la polarización de la onda incidente no coincide con la polarización de la antena receptora, tiene lugar una pérdida de polarización por desacoplamiento, que siempre debe ser tomada en cuenta en el diseño de los cálculos del enlace, especialmente en aplicaciones limitadas en cuanto a potencia.

Impedancia de entrada. Una antena debe estar conectada a un transmisor por medio de una línea de transmisión o guía de ondas, a fin de ser excitada y producir radiación. La impedancia de entrada de la antena presentada a la línea de alimentación constituye un parámetro importante, cuyo valor es necesario para el diseño de redes (circuitos) de acoplamiento eficientes que aseguren la máxima transferencia de potencia. La impedancia de entrada de la antena tiene en general una componente resistiva y una componente reactiva:

$$Z_A = R_A + jX_A \quad (48-102)$$

La componente reactiva se debe a los campos de inducción de la región próxima, debido a que tales campos producen un almacenamiento de energía reactiva en la región que rodea a la antena. La componente resistiva de la impedancia de entrada tiene contribuciones de todos los diversos elementos que provocan una pérdida de energía de la antena. En el caso de una antena en el espacio libre, donde no existe acoplamiento mutuo de otras fuentes, la resistencia de la antena puede concebirse como la suma de la resistencia de radiación R_r y la resistencia óhmica R_L :

$$R_A = R_r + R_L \quad (48-103)$$

La **resistencia de radiación** se define como la resistencia equivalente que disiparía una potencia igual a la potencia radiada cuando la corriente a través de la resistencia es igual a la corriente en las terminales de entrada de la antena. La resistencia óhmica explica las pérdidas debidas a una conductividad finita en la estructura de la antena. Para una antena eficaz, la resistencia de radiación debe ser mucho mayor que la resistencia óhmica. Por ejemplo, un dipolo delgado práctico de media onda tiene una resistencia de radiación de 73 Ω y una resistencia óhmica de unos 2 Ω .

La medición de la impedancia de entrada a altas frecuencias suele efectuarse midiendo el coeficiente de reflexión y la razón de voltajes de ondas estacionarias (VSWR, de *voltage standing wave ratio*). Esta última se relaciona con la magnitud del coeficiente de reflexión Γ por medio de:

$$\text{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \quad (48-104)$$

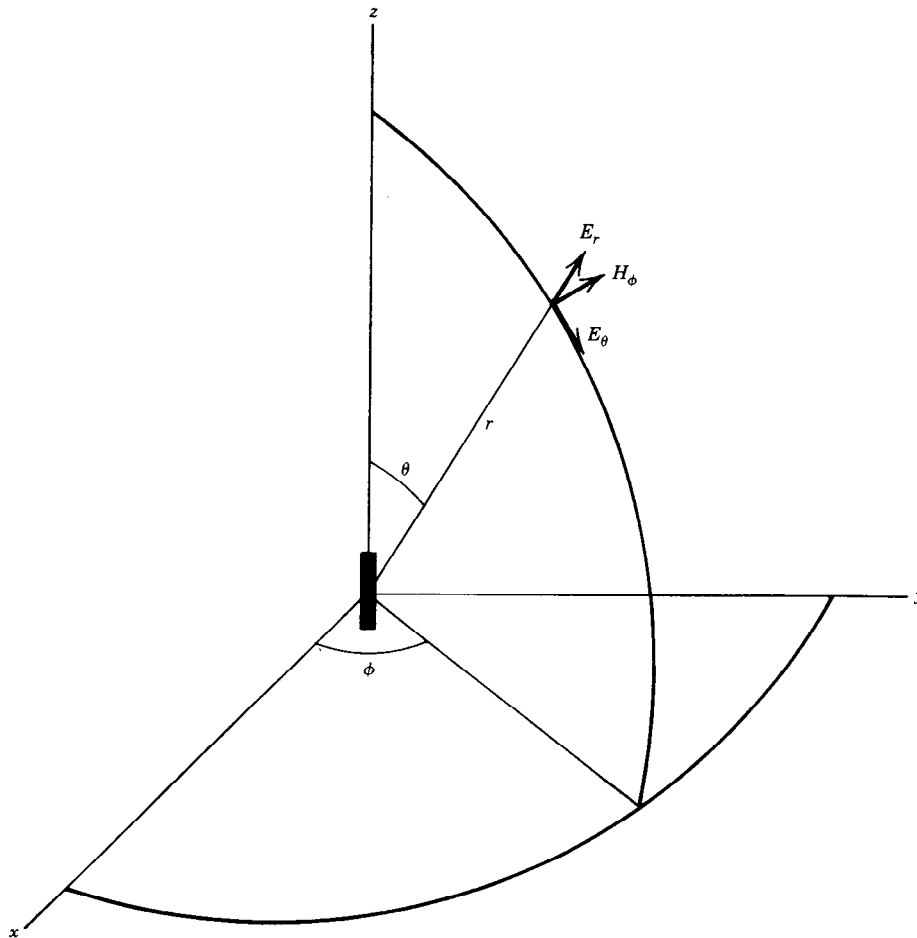


Fig. 48-68. Cálculo de los campos radiados de un dipolo corto.

La impedancia de la antena, Z_A , está dada por:

$$Z_A = Z_0 \left(\frac{1 + \Gamma}{1 - \Gamma} \right) \quad (48-105)$$

donde Z_0 es la impedancia característica de la línea de transmisión.

Antenas con elemento de corriente

Estas antenas son los radiadores de mayor uso en la actualidad, y pueden ser del tipo de corriente eléctrica, como las de dipolo y cuadro, o del tipo de corriente magnética, como la antena de ranura. El análisis se limitará aquí a los dipolos de longitud de onda corta y de longitud de onda media.

Considérese un dipolo corto, infinitamente delgado, colocado a lo largo del eje z , como se muestra en la figura 48-68. En la región de campo lejano, la radiación consta de una onda electromagnética transversal que se propaga alejándose del dipolo, y que está dada por:

$$E_\theta = j60\pi \frac{I\Delta z}{r\lambda} e^{-jKr} \sin\theta \quad (48-106)$$

$$H_\phi = \frac{E_\theta}{\eta_0}$$

donde I = corriente constante en el dipolo, A

Δz = longitud del dipolo, m

λ = longitud de onda, m

$K = 2\pi/\lambda$ = constante de propagación, rad/m

$j = \sqrt{-1}$

$\eta_0 = 377$ ohms = impedancia del espacio libre

Si la longitud del dipolo se incrementa a la mitad de una longitud de onda ($\lambda/2$), la distribución de corriente deja de ser uniforme, aunque en una primera aproximación es igual a $I(Z) = I_0 \cos KZ$. Entonces, el campo lejano está dado por:

$$E_\theta = j60I_0 \frac{e^{-jKr}}{r} \frac{\cos[(\pi/2)\cos\theta]}{\sin\theta} \quad (48-107)$$

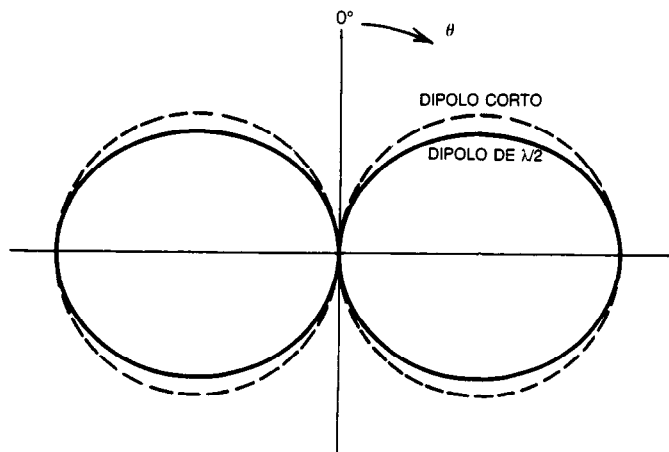


Fig. 48-69. Diagrama en alzado de los patrones de un dipolo vertical.

En la figura 48-69 se comparan los diagramas de proyección vertical (en alzado) del dipolo corto y el de $\lambda/2$. En el plano acimutal (plano xy), la radiación de cualquier antena es omnidireccional y el patrón una circunferencia.






Antenas de abertura

Existe un gran número de tipos de antenas para los cuales el campo electromagnético irradiado puede considerarse como si emanara de una abertura física. Entre las antenas comprendidas en esta clase se incluyen varios tipos de reflectores, lentes y bocinas, por mencionar unas cuantas⁶. En general, la abertura es un orificio finito en un plano infinito. El campo eléctrico, magnético o ambos en la región de la abertura se de-

terminan en primer lugar por medio de métodos de aproximación; después se calculan los campos radiados, utilizando como fuentes solamente los campos en la superficie de la abertura.

En la tabla 48-9 se enumeran las características de los patrones de una abertura circular de diámetro D en distintas distribuciones de ahusamiento. Para cada distribución de abertura se proporcionan el ancho del haz de 3 dB (en radianes), el nivel del lóbulo lateral y la posición del primer cero de la característica. La distribución uniforme tiene una característica de campo lejano de $\sin x/x$ y nivel del lóbulo lateral de -13 dB. Las distribuciones ahusadas reducen el nivel del lóbulo lateral a expensas de algo de ensanchamiento del haz y de reducción en la ganancia. La distribución de Taylor

Tabla 48-9. Distribuciones de la abertura circular

Distribución	Campo de abertura	Amplitud del haz 3 dB	Nivel del primer lóbulo lateral	Posición angular del primer cero
$0 \leq r \leq 1$				
Uniforme		$1.02 \frac{\lambda}{D}$	-17.6 dB	$1.22 \frac{\lambda}{D}$
Ahusada a cero en el borde ($1 - r^2$)		$1.27 \frac{\lambda}{D}$	-24.6 dB	$1.63 \frac{\lambda}{D}$
Ahusada a cero en el borde ($1 - r^2$) ²		$1.47 \frac{\lambda}{D}$	-30.6 dB	$2.03 \frac{\lambda}{D}$
Ahusada a 0.5 en el borde [$0.5 + (1 - r^2)^2$]		$1.16 \frac{\lambda}{D}$	-26.5 dB	$1.51 \frac{\lambda}{D}$
Distribución de Taylor		$1.31 \frac{\lambda}{D}$	-40.0 dB	

Fuente. Tomado de Rudge y colaboradores⁴, con permiso.

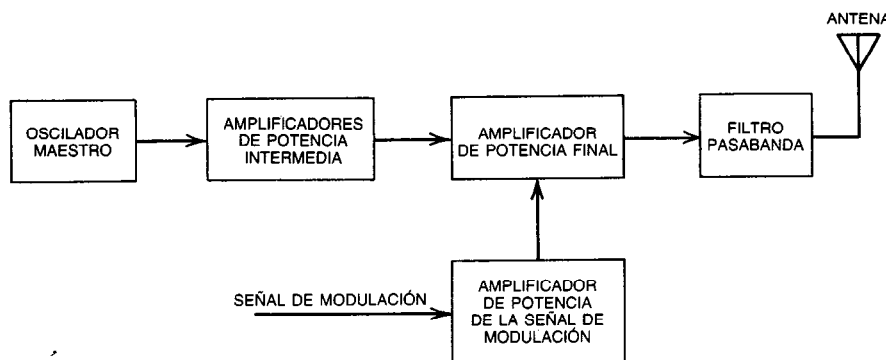


Fig. 48-70. Diagrama de bloques de un transmisor de AM.

es una distribución optimizada en la cual el nivel del lóbulo lateral es bajo y el haz principal sigue siendo razonablemente estrecho.

Redes de antenas

Estas antenas están constituidas por un grupo de elementos radiantes dispuestos en algún arreglo geométrico y eléctrico. La característica total del arreglo (también llamado sistema o red) se obtiene multiplicando la característica de un elemento individual (factor del elemento) y la característica debida a un sistema de fuentes puntuales isotrópicas (factor del sistema). El factor del sistema depende, por lo general, de su orientación geométrica, la separación entre elementos y la excitación de éstos. Usualmente los elementos radiantes individuales tienen bajas directividades, por lo cual la característica del sistema está determinada en gran medida por el factor del propio sistema, aunque puede haber excepciones para esta regla. Algunos elementos típicos de los sistemas son los dipolos, monopolos, guías de ranura de onda, guías de ondas de extremo abierto y radiadores de microbanda. La elección de un elemento específico para una aplicación en particular depende de varios factores, tales como la frecuencia, capacidad de manejo de potencia, polarización, configuración de alimentación, etc. El arreglo (red) se denomina lineal cuando los elementos se encuentran en línea recta, planar cuando se encuentran en un plano, y conforme cuando los elementos se apegan a alguna configuración geométrica como una circunferencia o una esfera.

Una de las características atractivas de un arreglo es la capacidad de barrer electrónicamente el haz radiado mediante el control de la fase de excitación de los elementos individuales. Cuando se utiliza de esta forma, el arreglo se denomina red de elementos en fase. Dos sistemas avanzados de considerable interés actualmente son las redes para el procesamiento de señales y las redes autoadaptativas. En estas últimas, la característica de radiación del sistema se adapta por sí misma a alguna situación específica bajo el control de una computadora. En particular, la red de orientación de ceros dirige los ceros en la característica de radiación de modo que coincidan con la dirección de alguna señal indeseada o de interferencia.

48.2.4 Fundamentos sobre transmisores

Introducción

Los transmisores pueden clasificarse con facilidad con arreglo a la técnica de modulación en la que se basan. Entre estas técnicas se incluyen las de modulación de amplitud (AM), modulación de frecuencia (FM), modulación de fase (PM) y banda lateral única (SSB). La elección de un tipo dado de transmisor suele depender de la aplicación que se pretenda, la frecuencia de operación y las diversas reglas y reglamentaciones que rigen la radiotransmisión, como las promulgadas por la FCC, el CCITT, el ITU y otras organizaciones internacionales.

En esta subsección se describen, a nivel de diagramas de bloques, los tipos más elementales de transmisores. En la sección 48.1 se explica brevemente la teoría de la modulación relacionada con cada tipo de transmisor.

Transmisor de AM

La estructura acostumbrada de este tipo de transmisor es un oscilador maestro seguido de amplificadores de potencia, multiplicadores de frecuencia o ambos, en cantidad suficiente para obtener la salida de frecuencia y potencia deseada. La salida de esta etapa final se hace pasar por un filtro pasabajas a fin de atenuar armónicos, y después se envía a la antena. La modulación puede aplicarse a cualquiera de los amplificadores de potencia.

Dependiendo de la etapa de aplicación, los transmisores de AM se clasifican en dos grupos: los que emplean modulación de alto nivel y los que utilizan modulación de bajo nivel. La figura 48-70 es un diagrama de bloques de un transmisor común de AM de alto nivel, en el que la modulación se aplica al amplificador de potencia final. El oscilador maestro se utiliza para generar la frecuencia de la portadora. Sin embargo, si la frecuencia de operación es alta, se obtiene mejor estabilidad diseñando el oscilador para alguna frecuencia menor y luego utilizando multiplicadores de potencia para obtener el valor deseado de la frecuencia. Cuando se requiere capacidad para cambiar la frecuencia de operación, es posible utilizar las salidas de un oscilador de cristal y de un sintetizador de frecuen-

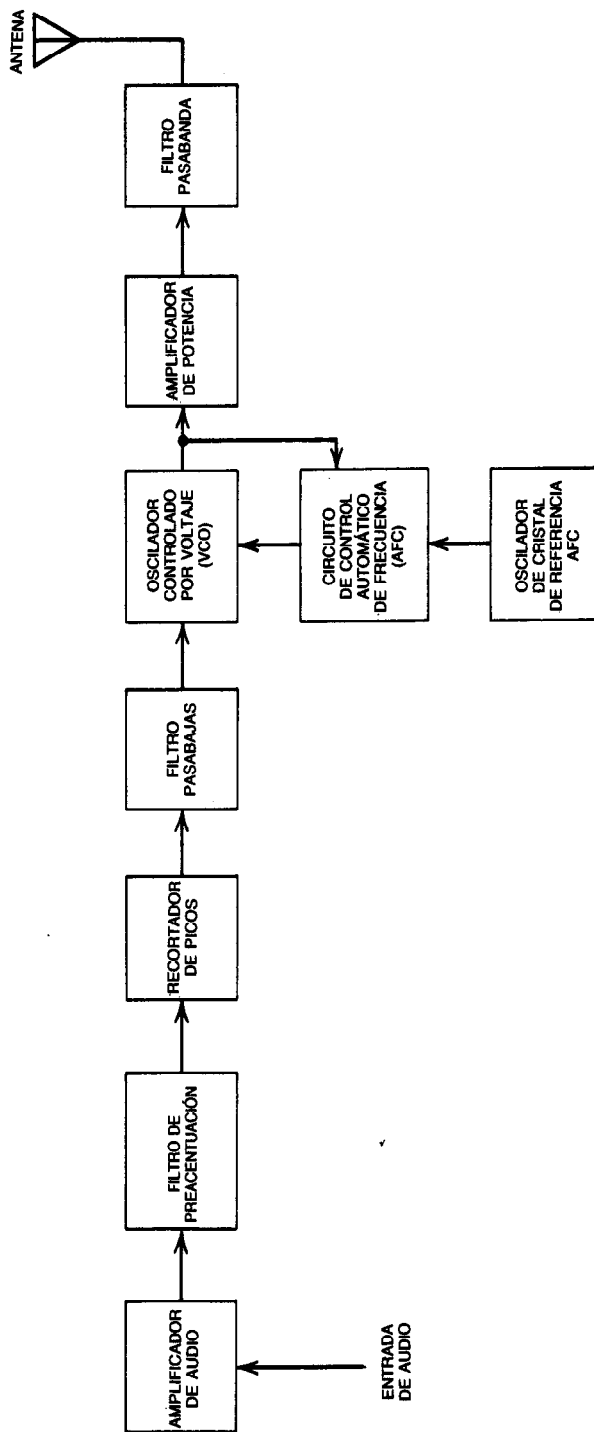


Fig. 48-71. Diagrama de bloques de un transmisor de FM común, en el que se utiliza el concepto de modulación directa.

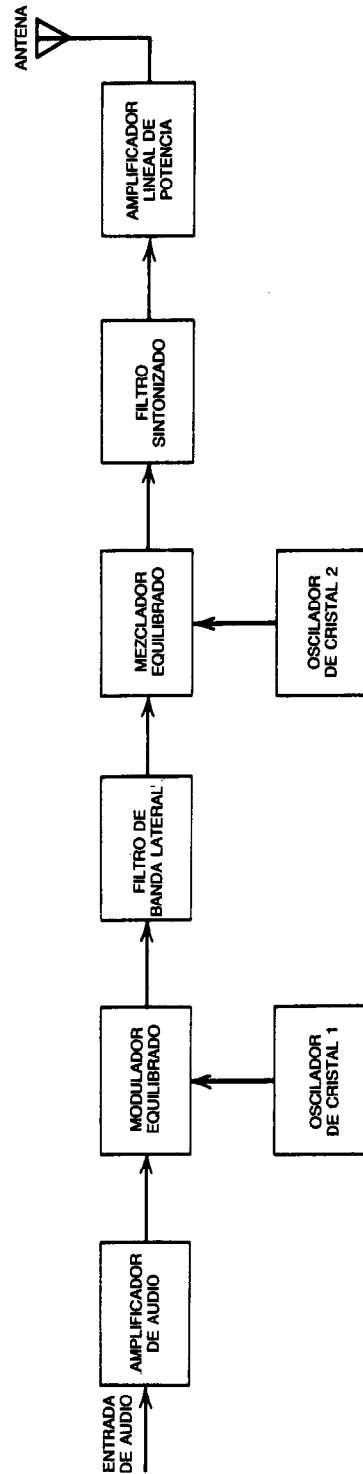


Fig. 48-72. Diagrama de bloques de un transmisor de banda lateral única (SSB) del tipo filtro.

cías como entradas a un mezclador cuya portadora de salida se transforme en una función de la entrada de control digital al sintetizador.

La ventaja de la modulación de alto nivel reside en el hecho de que todos los amplificadores de potencia intermedios pueden operarse en clase C a fin de obtener la máxima eficiencia. Esto es posible ya que la señal no está modulada. El número de etapas empleadas depende de la potencia necesaria para activar el amplificador de potencia final.

Una desventaja de la modulación de alto nivel es que para producirla se requiere una gran potencia de la señal moduladora.

En un transmisor de modulación de bajo nivel, la señal moduladora se aplica a uno de los amplificadores de potencia intermedios. Cuanto más temprana sea la etapa y más bajo sea el nivel de potencia de la portadora, menores serán los requisitos de potencia de la señal moduladora. Sin embargo, dado que ahora la señal de RF está modulada, todas las etapas ulteriores deben manejar potencia de banda lateral así como potencia de la portadora, y deben contar con ancho de banda suficiente para las frecuencias de la banda lateral. Además, todas estas etapas deben operarse en una clase lineal, A o B, con una pérdida acompañante en eficiencia⁷.

Transmisor de FM

Estos transmisores pueden clasificarse en dos grandes categorías, dependiendo de la técnica que se utilice para obtener la desviación de frecuencia. En la técnica de "modulación directa", la frecuencia del oscilador se hace variar en concordancia con la señal de modulación. En la técnica de "modulación indirecta", la modulación de la frecuencia se obtiene indirectamente después de modular la fase de la portadora.

En cada una de estas categorías, la FM puede subclasificarse en términos de la desviación de la frecuencia de cresta, ya sea como de "banda ancha" o de "banda estrecha". La FM de banda ancha, con desviación de cresta de ± 75 kHz, se utiliza cuando deben transmitirse señales de alta fidelidad, tal como en la radiotransmisión por FM y en el sonido para TV. La FM de banda estrecha, con desviación de cresta de ± 15 kHz, se utiliza en los denominados servicios móviles de comunicación por FM, como las radios de la policía, ambulancias, radiotaxis, etcétera.

En la figura 48-71 se muestra un transmisor de FM típico basado en el concepto de la modulación directa. El filtro de preacentuación o preénfasis, que precede al limitador de crestas, obliga a que primero se limiten las frecuencias más elevadas de audio, proporcionando de esta forma la misma desviación de cresta para todas las frecuencias de audio. El oscilador controlado por voltaje (VCO, de *voltage-controlled oscillator*), que está sintonizado ("enclavado") en fase con el oscilador de cristal con control automático de frecuencia (AFC, de *automatic frequency control*), genera la portadora modulada de la FM. El filtro pasabajas, ubicado entre el amplificador de potencia y la antena, atenúa cualquier señal espuria que pudiese resultar del acoplamiento con las señales de otros transmisores en las cercanías.

La configuración del transmisor de modulación en fase (PM) es parecida a la del transmisor de FM, con dos diferencias específicas. En el transmisor de PM no se utiliza el filtro de preénfasis, y la señal de modulación filtrada se aplica al circuito VCO de tal modo que se hace variar la fase del oscilador, en vez de la frecuencia. El transmisor de FM que se muestra en la figura 48-71 podría convertirse (conceptualmente) en un transmisor de PM sustituyendo el filtro de preénfasis por un diferenciador.

Transmisor de SSB

Básicamente, existen tres métodos prácticos para generar bandas laterales únicas: el método del filtro, el método del defasamiento y el método de Weaver. En los tres se utiliza el modulador equilibrado para suprimir la portadora, pero en cada uno se lleva a cabo una técnica diferente para eliminar la banda lateral no deseada.

Método del filtro. Es el más sencillo de los tres métodos, y en él la banda lateral indeseada se elimina por medio de un filtro pasabajas localizado a la salida del modulador equilibrado. El filtro puede ser de LC, de cristal o mecánico; este último es quizá el más popular. En la figura 48-72 se muestra un diagrama de bloques de un transmisor de SSB del tipo filtro. El mezclador equilibrado de la figura sirve para dos propósitos: por una parte, eleva la frecuencia a un valor descado para la transmisión; por otra parte, hace posible recuperar la suma de frecuencia (de la salida del filtro de la banda lateral y del oscilador núm. 2) con un filtro sintonizado.

Método del defasamiento. En éste se utilizan dos moduladores equilibrados y dos redes defasadoras, como se muestra en la figura 48-73. Un modulador recibe el voltaje del oscilador (defasado 90°) y el voltaje de modulación, mientras que el otro modulador recibe el voltaje de modulación (defasado 90°) y el voltaje del oscilador. Ambos moduladores producen una salida que consta solamente de bandas laterales. Sin embargo, con referencia al voltaje del oscilador, ambas bandas laterales superiores se adelantan 90° a la referencia, mientras que solamente una de las bandas laterales inferiores hace esto, y la otra se atrasa 90° . Por tanto, las dos bandas laterales inferiores están defasadas 180° entre sí, y cuando se suman en el combinador se cancelan mutuamente. Las bandas laterales superiores se encuentran en fase, por lo que se sumarán en el combinador, dando por resultado una señal SSB en la que se ha cancelado la banda lateral inferior. Si se desea generar la banda lateral inferior, un modulador debe tener ambas entradas defasadas 90° , mientras que el otro modulador debe tener sólo entradas no defasadas.

Método de Weaver. Es una extensión del método de defasamiento. Con él se elimina el problemático requisito de un defasador de audiofrecuencia de banda ancha y se sustituye por dos moduladores equilibrados extra, un generador de tono de audio y un solo defasador de audiofrecuencia¹.

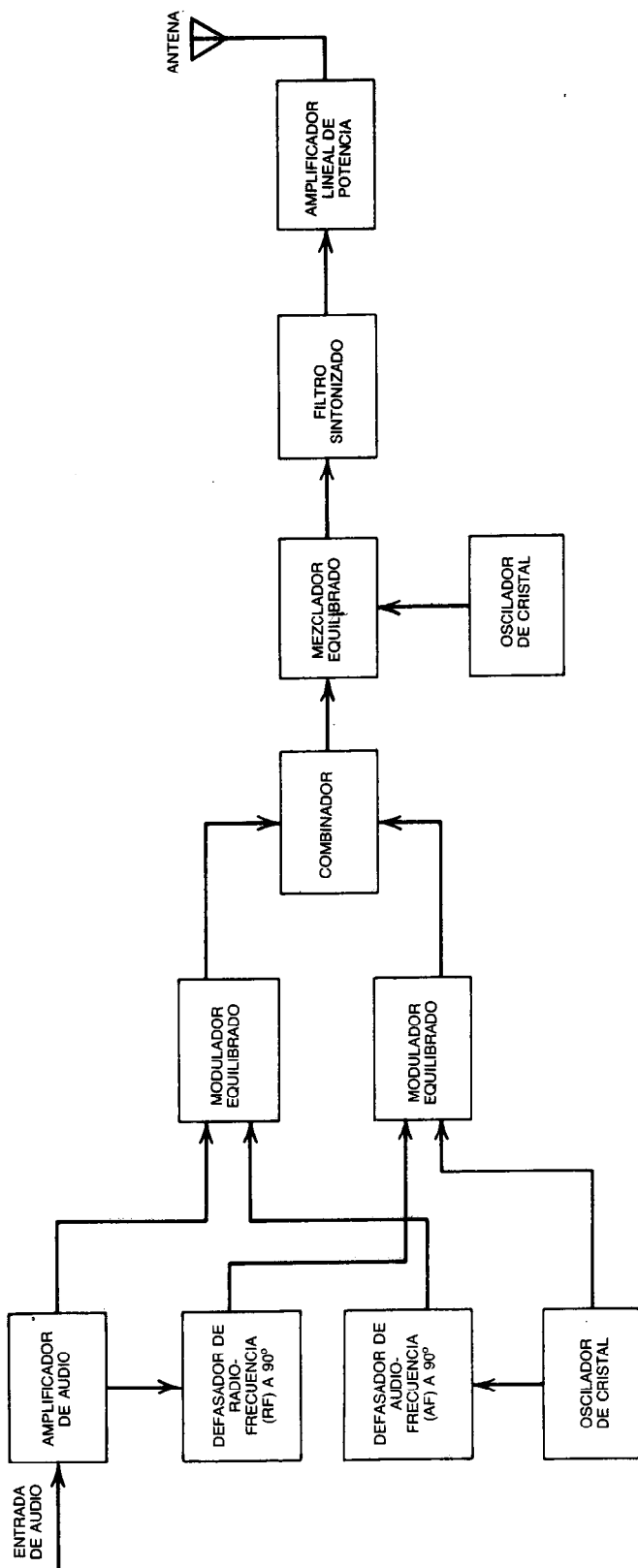


Fig. 48-73. Diagrama de bloques de un transmisor de banda lateral única (SSB) por defasamiento.

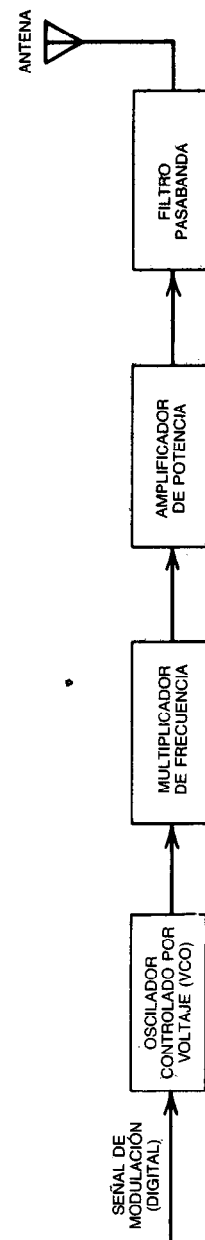


Fig. 48-74. Diagrama de bloques de un transmisor por desplazamiento de frecuencias (FSK).

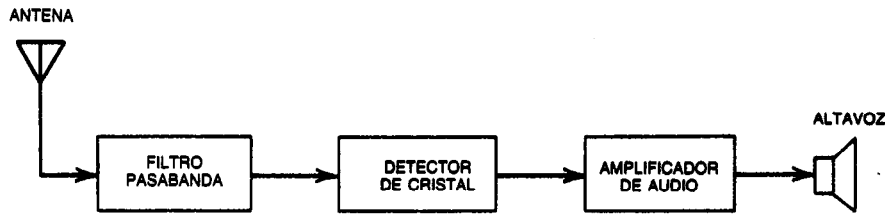


Fig. 48-75. Diagrama de bloques de un receptor de detector de cristal.

Transmisor de FSK

Uno de los más conocidos esquemas de modulación que se utilizan actualmente para la radiotelegrafía es el de manipulación por desplazamiento de frecuencias (FSK, de *frequency shift keying*), que puede concebirse como un sistema de modulación de frecuencia en el cual la frecuencia de la portadora se encuentra a mitad de camino entre la frecuencia de trabajo y la de reposo (también llamada de espacio o de pausa), y en el que la onda de forma de la modulación es una señal digital. Como resultado, el diagrama de bloques de un transmisor FSK es una versión abreviada del diagrama del transmisor de FM normal (fig. 48-74).

48.2.5 Fundamentos sobre receptores

Clasificación de los receptores

Un receptor es cualquier dispositivo que acepte y desmodule una señal de RF, a fin de extraer la información transmitida. Dado que el nivel de la señal a la entrada del receptor suele ser extremadamente bajo, es necesario amplificarlo en el receptor. La amplificación puede realizarse antes o después (o ambas cosas) de que se haya desmodulado la señal.

Los receptores suelen clasificarse según la modulación de la señal que deben recibir y desmodular. Por consiguiente, existen receptores de AM, FM, PM y varios otros.

Además, dentro de un esquema de modulación como la AM, los receptores pueden subdividirse conforme al tipo de equipo utilizado en la estructura del receptor. En este sentido es posible establecer una diferencia entre los receptores de detector de cristal, de RF sintonizada, superheterodinos, regenerativos y superregenerativos.

Receptor de detector de cristal. En la figura 48-75 se muestra el diagrama de bloques de un receptor de detector de cristal. El filtro pasabanda, localizado a la entrada del detector, proporciona cierta selectividad al receptor. El detector de cristal es un dispositivo de señal débil o de ley cuadrática, y toda la amplificación en el receptor es proporcionada por el amplificador de audio. La única ventaja de este tipo de receptor es la sencillez de su circuitería, ya que no hay oscilador local, amplificadores sintonizados, ni detector de ley cuadrática. Algunas de sus desventajas son sensibilidad deficiente y el requisito de una elevada amplificación de audio.

Receptor de radiofrecuencia sintonizada (TRF, de *tuned radio-frequency*). En la figura 48-76 se presenta el diagrama de bloques del receptor TRF. Varios amplificadores de RF dispuestos en cascada, todos sintonizados a la misma frecuencia, amplifican la señal recibida a un nivel adecuado para la detección. Después, la señal de audio se detecta, amplifica y envía al altavoz. Generalmente sin pérdida, la señal detectada puede considerarse como una señal de vídeo o de telemetría que se envía a un monitor de imágenes o a un dispositivo de grabación.

El receptor TRF está caracterizado por varias ventajas y desventajas inherentes, en comparación con el actualmente más común receptor superheterodino. Dos ventajas que pueden mencionarse casi de forma instantánea son la sencillez de su diseño y la ausencia de problemas de frecuencia de la imagen (espectral). La lista de desventajas es mucho más grande:

1. El receptor TRF presenta una deficiente selectividad a altas frecuencias, debido al empleo obligado de circuitos resonantes simples (de un solo elemento sintonizado). La selectividad deficiente, a su vez, da por resultado interferencias del canal adyacente.
2. Para ganancia óptima todos los circuitos tanque de las diversas etapas de RF deben sintonizarse a la misma frecuencia en todo el intervalo de sintonización. Tal exactitud o "perfección" de "seguimiento" resulta muy difícil de lograr en la práctica. Por tanto, se obtiene una variación en la ganancia o sensibilidad como función de la frecuencia de sintonización.
3. El riesgo de inestabilidad es bastante real a alta frecuencia siempre que se obtiene elevada ganancia por medio de un amplificador de etapas múltiples. En tal caso, la retroalimentación positiva debida a capacitancia parásita o a trayectorias de dispersión (erráticas) puede hacer oscilar el amplificador e inutilizar el receptor.

Receptor superregenerativo. Es un amplificador o detector de RF que tiene suficiente retroalimentación positiva para provocar oscilación. El receptor se hace entrar y salir de la oscilación por medio de una señal de control, denominada señal de "amortiguamiento", cuyas frecuencias típicas varían de 10 kHz a 1 MHz. Con este tipo de receptor es posible obtener ganancias muy elevadas.

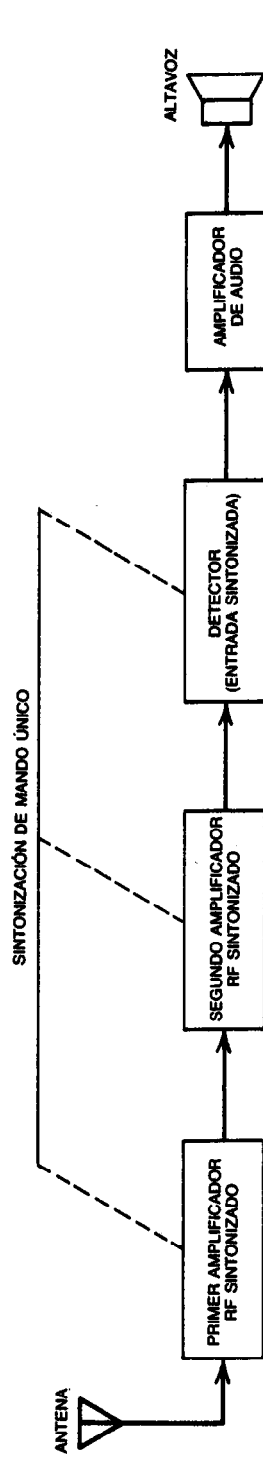


Fig. 48-76. Diagrama de bloques de un receptor de RF sintonizada.

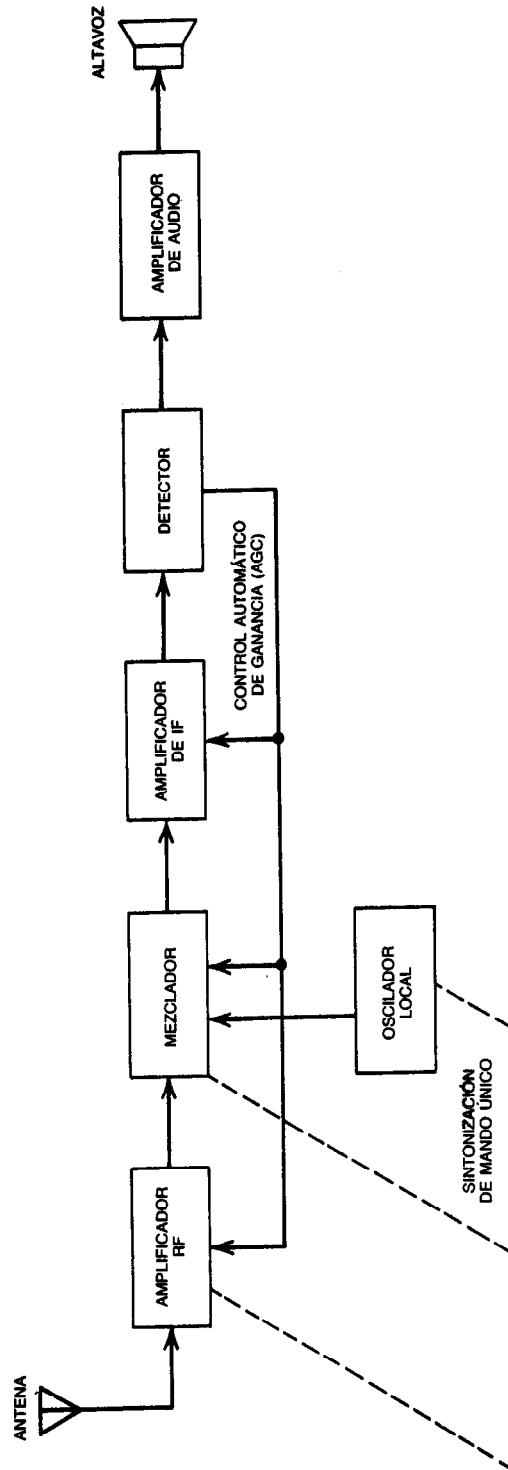


Fig. 48-77. Diagrama de bloques de un receptor superheterodino.

La radiación de los receptores superregenerativos es bastante alta, de modo que pueden presentarse problemas si es necesario que varios receptores funcionen muy cerca unos de otros. Este tipo de receptor no es muy apropiado para la recepción normal de varios canales, aunque puede servir bastante bien para pequeños radiofaros, transpondedores y aplicaciones de control remoto. Durante un ciclo de amortiguamiento se detecta una muestra de la señal. Una portadora modulada dará por resultado una señal de salida constante si la frecuencia amortiguadora es igual a la frecuencia de modulación. De otro modo, la frecuencia de salida es igual a la frecuencia de modulación menos la frecuencia amortiguadora⁸.

Receptor superheterodino. Prácticamente toda forma de radorreceptor en uso actualmente se basa en el concepto superheterodino, que puede encontrarse en aplicaciones tan variadas como radiocomunicación, televisión y receptores de radar, con solamente ligeras modificaciones en principio. Aunque el receptor superheterodino es más complicado que los demás tipos, en general presenta mayor sensibilidad y selectividad.

En la figura 48-77 se muestra el diagrama de bloques de un receptor superheterodino. El principio básico es la conversión de la señal RF a una frecuencia intermedia (IF) mediante la heterodinación de la señal RF con un oscilador local cuya frecuencia difiera de la propia de la señal RF por una cantidad igual a la frecuencia intermedia (IF) deseada. La sintonización del receptor consiste en la sintonización simultánea del oscilador local, el mezclador y el amplificador de RF, de modo que la frecuencia intermedia (IF) permanezca constante.

En el amplificador de IF suele utilizarse un gran número de circuitos de doble resonancia que funcionan a una frecuencia constante a fin de satisfacer la mayor parte de los requisitos de ganancia y ancho de banda del receptor. Como resultado de la menor frecuencia y la sintonización constante del amplificador de IF, la sensibilidad y la selectividad son esencialmente constantes en todo el intervalo de sintonización del receptor. La sección de RF se utiliza principalmente para atrapar la frecuencia deseada, rechazar la interferencia de la "frecuencia imagen" y reducir el factor de ruido del receptor.

Rendimiento del receptor

Sin tomar en cuenta la aplicación específica de un receptor, cuatro factores principales rigen su rendimiento: la sensibilidad, selectividad, fidelidad y factor de ruido. Estos factores dependen de la combinación de la amplificación, linealidad de fase y amplitud, respuesta de frecuencia, estabilidad y ruido en las diferentes etapas que integran el receptor.

Sensibilidad. La sensibilidad de un receptor de radio es su capacidad de amplificar señales débiles. Se define en términos del voltaje que debe aplicarse a las terminales de entrada del receptor a fin de obtener una potencia de salida normal, medida en las terminales de salida. Una sensibilidad de aproximadamente 50 μV es

típica en muchos receptores AM de banda amplia. Los receptores para una comunicación de calidad pueden presentar sensibilidades por debajo de 1 μV en la banda HF. Debe elegirse el nivel de sensibilidad mayor que el nivel de ruido generado dentro del receptor. En caso contrario la salida del receptor contendrá más ruido que otra cosa, con lo que se desperdiciará parte de la amplificación.

Selectividad. La selectividad de un receptor es su capacidad para rechazar todas las señales indeseadas, y es función de la respuesta de frecuencia de los circuitos sintonizados que se encuentran antes del detector. Gráficamente puede representarse como una curva que muestre el incremento necesario en el nivel de la señal para mantener la salida estándar de varias frecuencias de entrada sobre ambos lados de la frecuencia de resonancia del receptor (fig. 48-78). Debe tenerse cuidado de no hacer muy pronunciada la selectividad y con ello amenazar la integridad de la energía de la banda lateral de la señal deseada.

En términos generales, la selectividad está determinada por la respuesta de la sección de IF (frecuencia intermedia); la sección RF tiene un cometido pequeño, pero importante. La sección RF determina la selectividad de entrada (o selectividad de RF) del receptor, que es necesaria para el rechazo de la "frecuencia imagen". Esta frecuencia está dada por:

$$f_{s1} = f_s + 2f_i \quad (48-108)$$

donde f_s es la frecuencia deseada de la señal y f_i es la frecuencia intermedia, IF. Si una señal de interferencia de frecuencia f_{s1} no es verificada por medio de la porción de RF, producirá una respuesta espuria en la etapa de IF y comprometerá el funcionamiento del receptor.

Fidelidad. Ésta es la que indica cuán exactamente el receptor reproduce la señal de información original. La pérdida de fidelidad puede atribuirse a no linealidades en las diferentes etapas del receptor, que pueden producir efectos indeseados, como:

1. Modulación indeseada.
2. Distorsión de la modulación.
3. Modulación cruzada.
4. Generación de señales espurias.

Factor de ruido. Un receptor "ideal" es aquel que no presenta ruido generado internamente. El factor de ruido de un receptor práctico es una medida del ruido que él produce con respecto al de un receptor ideal. Este factor puede definirse como:

$$F = \frac{S_{\text{ent}}/N_{\text{ent}}}{S_{\text{sal}}/N_{\text{sal}}} \quad (48-109)$$

donde S_{ent} = potencia disponible de la señal de entrada

N_{ent} = potencia disponible del ruido de entrada

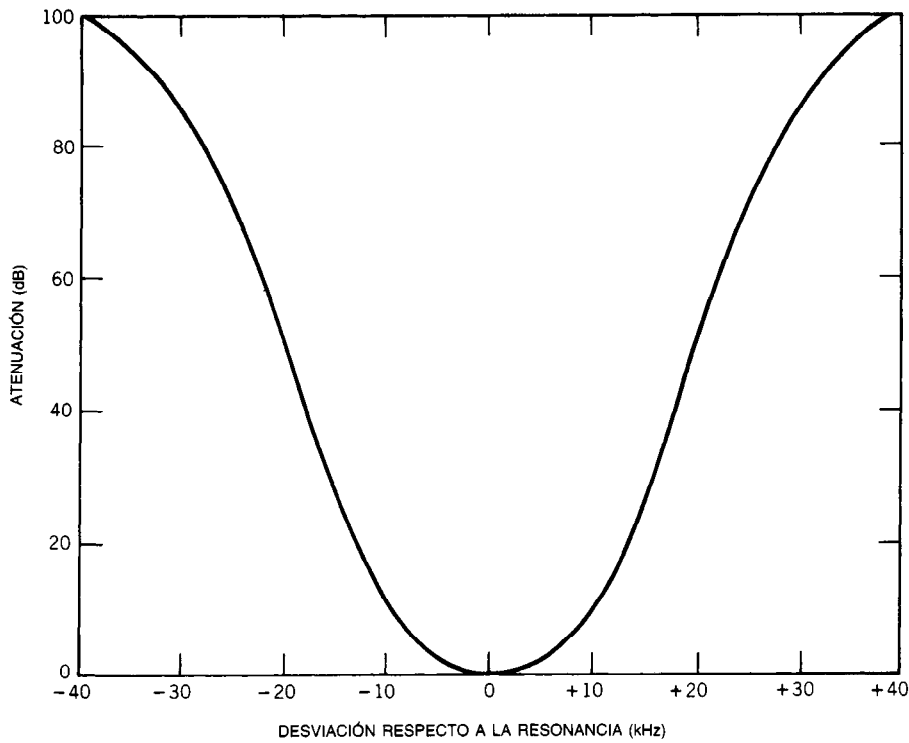


Fig. 48-78. Curva de selectividad típica.

S_{sal} = potencia disponible de la señal de salida

N_{sal} = potencia disponible del ruido de salida

El término "potencia disponible" se refiere a la potencia que se entregaría a una carga acoplada. La potencia del ruido en la entrada del receptor, N_{ent} , está dada por:

$$N_{ent} = KT_0B \text{ watts} \quad (48-110)$$

donde $K = 1.38 \times 10^{-23} \text{ J/K}^\circ$ = constante de Boltzmann

$T_0 = 290 \text{ K}$ = temperatura ambiente

B = ancho de banda del ruido

Utilizando la definición de ganancia del receptor $G = S_{sal}/S_{ent}$ y sustituyendo la potencia del ruido en la entrada N_{ent} de la ecuación 48-110, en la ecuación 48-109 se obtiene:

$$F = \frac{N_{sal}}{KT_0BG} \quad (48-111)$$

Con base en la ecuación 48-111, el factor de ruido puede interpretarse como la razón de la verdadera potencia disponible del ruido de salida entre la potencia del ruido que se tendría disponible si el receptor no tuviese ruido generado internamente. El factor de ruido también puede representarse como:

$$F = 1 + \frac{\Delta N}{KT_0BG} \quad (48-112)$$

donde ΔN es la potencia de ruido introducida por el receptor en sí. La temperatura de ruido efectiva del receptor se define como la temperatura T_e a la entrada del receptor, que explicaría el ruido agregado ΔN a la salida. Por consiguiente:

$$\Delta N = KT_eBG$$

y:

$$F = 1 + \frac{T_e}{T_0} \quad (48-113)$$

o bien:

$$T_e = (F - 1)T_0 \quad (48-114)$$

Para un receptor ideal, $F = 1$ (0 dB); $T_e = 0 \text{ K}$.
Para $F = 2$ (3 dB); $T_e = 290 \text{ K}$.

Receptores de AM

Estos receptores, que se utilizan en la actualidad, tienen básicamente la misma estructura que el receptor superheterodino recientemente descrito. Ciertos aspectos del diagrama de bloques de la figura 48-77 se considerarán ahora en el contexto de la modulación de amplitud.

En los receptores AM de la radiodifusión estándar, el intervalo de sintonización de la portadora varía de 540 a 1 650 kHz, con una frecuencia intermedia cons-

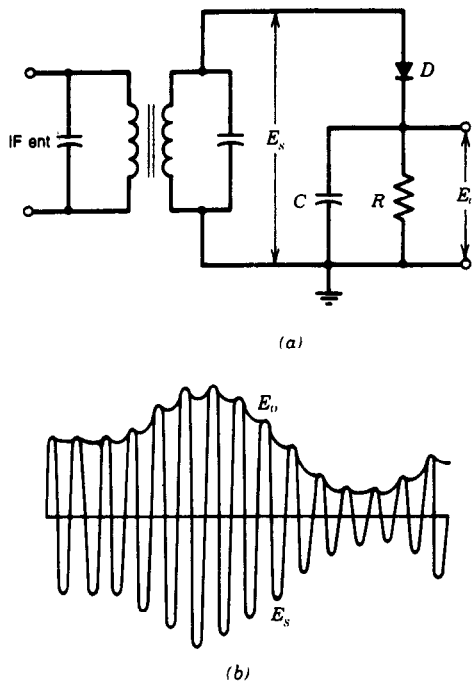


Fig. 48-79. Detector de envoltura simple: a) diagrama circuital; b) voltajes de entrada y de salida.

tante de 455 kHz. Para el caso usual de frecuencia del oscilador local por encima de la frecuencia de la portadora, el intervalo correspondiente de frecuencias del oscilador local es de 995 a 2 105 kHz, lo que da una razón de frecuencias máxima entre mínima de 2.2:1, que puede obtenerse mediante sintonización por capacitor variable.

Para recuperar la señal de información, en un receptor de AM se utiliza un detector de envoltura, que en su forma más sencilla tiene el diagrama circuital que se muestra en la figura 48-79. La señal rectificada de media onda en la salida del diodo se carga por medio del circuito en paralelo RC . En cada cresta positiva del ciclo de RF, C se carga hasta el voltaje de cresta de la señal E_s . En el breve lapso de tiempo entre las crestas RF, el capacitor C se descarga ligeramente en el resistor R , sólo para ser recargado en la siguiente cresta positiva. El resultado es el voltaje E_o , que reproduce con exactitud el voltaje de modulación, excepto en la pequeña cantidad de la fluctuación RF. Obsérvese que la constante de tiempo RC debe ser lo suficientemente grande para mantener pequeña la fluctuación RF, y lo suficientemente pequeña para seguir las rápidas variaciones en la señal de modulación. Los detectores de envoltura que se utilizan en la práctica presentan características que eliminan la componente de CD y la pequeña fluctuación RF de la onda de forma detectada.

Otro aspecto del diseño de receptores que debe mencionarse en este punto es el concepto de control de ganancia automático (AGC, de *automatic gain con-*

trol). El AGC, como suele denominarse, es una técnica por medio de la cual la ganancia global de un receptor de radio se hace variar automáticamente con la intensidad cambiante de la señal recibida, a fin de mantener la salida sustancialmente constante. Esto se logra mediante un voltaje de polarización CD, deducido por el detector, y aplicado a un número selecto de etapas de RF, IF y mezclador. El AGC permite la sintonización de estaciones que presentan gran disparidad en las intensidades de la señal portadora sin un cambio apreciable en el nivel de la señal de salida. También ayuda a suavizar el rápido desvanecimiento que podría ocurrir en la recepción de onda corta a larga distancia, y evita la sobrecarga de la última etapa de amplificador de IF en el caso de señales recibidas excesivamente fuertes.

Receptores de FM

Los canales de comunicación por FM se encuentran a lo largo de todo el espectro de frecuencias. Sin importar la frecuencia, un receptor de FM tiene una estructura superheterodina semejante a la de su contraparte de AM. Como puede observarse en el diagrama de bloques de la figura 48-80, existen algunas diferencias entre el receptor de AM y el de FM. Algunas características únicas de este último son: limitación en amplitud, desacentuación, una técnica de modulación diferente y un método también diferente para generar el AGC. Por lo general, la IF elegida es una función de la banda de la frecuencia portadora y de la clase de servicio de comunicación. Por ejemplo, los receptores de radiodifusión estándar que operan en el intervalo de frecuencias de 88 a 108 MHz utilizan una IF de 10.7 MHz.

Todos los desmoduladores de FM están precedidos por un limitador de amplitud que elimina las variaciones de la amplitud en la señal de IF. Si estas variaciones no se eliminan, pueden producir distorsión en la salida del desmodulador.

La función principal de un desmodulador de FM es como convertidor de frecuencia a amplitud. Traduce la desviación de la frecuencia de la señal de IF de entrada en una variación de amplitud que es una reproducción de la señal de información. Tres de los más conocidos desmoduladores de FM son: el detector de pendiente equilibrado, el discriminador de Foster-Seeley y el detector de razón.

El detector de pendiente equilibrado divide la señal de FM en dos componentes defasadas 180° y envía cada una a un circuito sintonizado separado. Los dos circuitos sintonizados tienen resonancias por encima y por debajo de la frecuencia central, respectivamente. Sus salidas se hacen pasar por detectores a diodo separados, y luego se suman entre sí a fin de obtener la salida final del desmodulador.

En el discriminador de Foster-Seeley se utilizan dos circuitos sintonizados que son resonantes en la frecuencia central. En este caso la desviación de la frecuencia se traduce a una diferencia de fase entre las salidas de los dos circuitos sintonizados. Debido a su dependencia respecto a las relaciones de fase, este desmodulador también se denomina discriminador de fa-

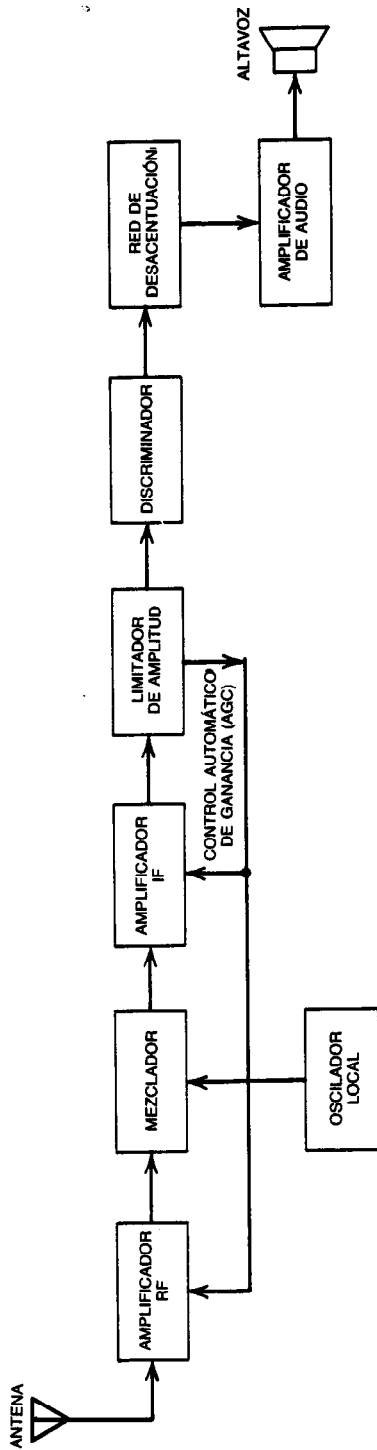


Fig. 48-80. Diagrama de bloques de un receptor de FM.

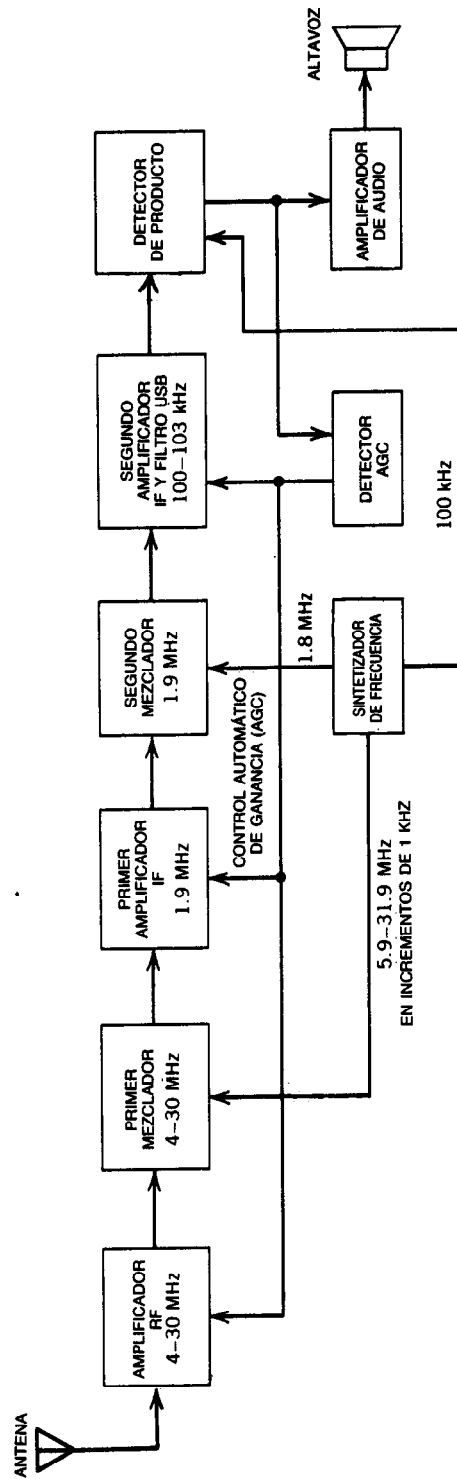


Fig. 48-81. Diagrama de bloques de un receptor de banda lateral única con supresión de la portadora (SSBSC).

se. Es mucho más fácil de alinear y proporciona mejor linealidad que el detector de pendiente equilibrado.

El detector de razón funciona con el mismo principio que el discriminador de Foster-Seeley, pero tiene la ventaja de contar con una capacidad inherente de limitación del voltaje, lo que hace innecesaria la limitación previa.

Receptores de SSB

Los receptores de banda lateral única (SSB, de *single sideband*) suelen ser necesarios para operar en bandas congestionadas, donde la atenuación de las interferencias de los canales adyacentes es de suma importancia. Por esta causa suelen ser de la variedad de doble conversión, en la que se incorporan dos mezcladores y se generan dos frecuencias intermedias. La primera IF, alta, empuja la frecuencia imagen, alejándola de la frecuencia de señal y permitiendo de esta forma mucha mejor atenuación. La segunda IF (baja), por otra parte, presenta todas las ventajas de una frecuencia de operación fija baja, selectividad especialmente aguda y por tanto buen rechazo del canal adyacente.

Para desmodular la señal SSB en la salida del segundo amplificador IF, en los receptores de SSB por lo general se utiliza uno de dos métodos: el del detector de producto o el del modulador equilibrado.

Detector de producto. Se parece a un mezclador ordinario en que se multiplica la señal SSB con la frecuencia portadora nominal o con la frecuencia portadora piloto, si ésta es transmitida. Luego se separa el audio de baja frecuencia de los términos de producto mediante un filtro pasabajas.

Modulador equilibrado. Suele utilizarse en transceptores, en los que puede emplearse tanto para modulación como para desmodulación mediante conmutación apropiada. Como en el detector de producto, se generan suma y diferencia de frecuencias entre la señal SSB y una portadora nominal, y luego se retienen y entregan a la salida sólo los términos de baja frecuencia, o de audio.

En la figura 48-81 se muestra el diagrama de bloques de un receptor de SSB con supresión de la portadora. El receptor de banda lateral superior utiliza un sintetizador de frecuencias a fin de generar osciladores locales estables para los dos mezcladores y el detector de producto. La entrada del oscilador al primer mezclador es en pasos de 1 kHz, lo que hace necesario que el transmisor se canalice a frecuencias de 1 kHz para una recepción correcta. El voltaje AGC se obtiene por rectificación parcial de la salida de audio.

El receptor de portadora piloto es otro tipo de receptor de SSB en el cual la portadora piloto recuperada sirve para dos propósitos: se utiliza como referencia para un circuito de control de frecuencia automático a fin de asegurar una buena estabilidad global de frecuencia, y, por otra parte, dado que la amplitud de la portadora piloto varía con la intensidad de la señal de entrada, es posible obtener de ella una señal AGC aceptable por simple rectificación.

48.2.6 Enlace de radiocomunicación total

Análisis del enlace de comunicación

El funcionamiento de extremo a extremo de cualquier enlace de comunicación puede expresarse en términos de la razón de señal sobre ruido (SNR) evaluada en un ancho de banda apropiado en algún punto del sistema de recepción. En un sistema analógico en el cual el ancho de banda del ruido suele ser mayor que el ancho de banda de la señal, a menudo se recurre a la razón promedio de potencia sobre ruido de la portadora, o (P_r/N) , como el SNR de interés particular:

$$\frac{P_r}{N} = \frac{\text{EIRP}(G_r/T_e)}{L_p L_{oe} L_{oi} KB} \quad (48-115)$$

Los diversos parámetros de la ecuación 48-115 pueden definirse con la ayuda del modelo de enlace que se muestra en la figura 48-82,

donde P_r = potencia de la señal recibida en la entrada del detector, W

$N = KT_e B$ = potencia del ruido térmico a la entrada del detector, W

K = constante de Boltzmann, 1.38×10^{-23} J/K

T_e = temperatura de ruido equivalente del sistema, K

B = ancho de banda a la entrada del detector, Hz

$\text{EIRP} = P_t G_t$ = potencia radiada isotrópicamente equivalente, W

P_t = potencia transmitida, W

G_t = ganancia de la antena de transmisión, adimensional

G_r = ganancia de la antena de recepción, adimensional

G_r/T_e = factor de mérito, razón de ganancia sobre temperatura de ruido equivalente del sistema, K^{-1}

$L_p = (4\pi d/\lambda)^2$ = pérdida de espacio libre, ya definida

L_{oe} = otras pérdidas externas = pérdida atmosférica + pérdida por polarización de antena + pérdida por direccionamiento de antena

L_{oi} = otras pérdidas internas = pérdida del circuito de transmisión + pérdida del circuito de recepción + pérdida por ruido de intermodulación

En un sistema digital en el que el ancho de banda de la señal se considera igual al ancho de banda del ruido, el rendimiento del enlace se expresa en términos de la razón potencia de señal recibida sobre densidad espectral del ruido (P_r/N_0):

$$\frac{P_r}{N_0} = \frac{\text{EIRP}(G_r/T_e)}{L_p L_{oe} L_{oi} KB} \quad (48-116)$$

donde $N_0 = N/B$ = densidad espectral del ruido en watts por hertz.

Si se supone que toda la potencia recibida proviene de la señal de modulación (portadora suprimida), en-

tonces es posible evaluar el rendimiento del enlace en términos de la razón de densidad espectral del ruido sobre energía en los bits. Puede expresarse:

$$\frac{P_r}{N_0} = \left(\frac{E_b}{N_0} \right) R \quad (48-117)$$

Entonces:

$$\frac{E_b}{N_0} = \frac{\text{EIRP}(G_r/T_e)}{L_p L_{oe} L_{oi} K R} \quad (48-118)$$

donde R es la velocidad de transmisión de datos de información en bits por segundo. Si la potencia de la portadora no es despreciable, aún puede utilizarse la ecuación 48-118, suponiendo que la potencia de la portadora se considera una pérdida en el parámetro L_{oi} . En decibels, la ecuación 48-118 puede expresarse como:

$$\begin{aligned} \frac{E_b}{N_0} \text{ (dB)} &= \text{EIRP(dBW)} + \frac{G_r}{T_e} \text{ (dB/K)} - \\ &- L_p \text{ (dB)} - L_{oe} \text{ (dB)} - L_{oi} \text{ (dB)} - \\ &- K \text{ (dBW/K - Hz)} - R \text{ (dB Hz)} \end{aligned} \quad (48-119)$$

El término E_b/N_0 , definido como en la ecuación 48-118, se refiere al E_b/N_0 necesario para el logro de una cierta probabilidad de errores en los bits, P_b , en la salida del detector. En la práctica, el E_b/N_0 real está relacionado con el E_b/N_0 requerido por un factor de seguridad M , que suele denominarse margen de enlace:

$$\left(\frac{E_b}{N_0} \right)_{\text{actual}} = M \left(\frac{E_b}{N_0} \right)_{\text{req}} \quad (48-120)$$

En decibels, el margen del enlace es sencillamente la diferencia entre los valores real y requerido de E_b/N_0 :

$$M \text{ (dB)} = \left. \frac{E_b}{N_0} \right|_{\text{actual}} \text{ (dB)} - \left. \frac{E_b}{N_0} \right|_{\text{req}} \text{ (dB)} \quad (48-121)$$

Para una probabilidad dada de errores en los bits, el E_b/N_0 necesario es función de las técnicas de modulación y codificación utilizadas en el enlace. La cantidad de margen de enlace por utilizar depende en gran medida de la naturaleza estadística del enlace y de la habilidad que se tenga para predecir todas las fuentes de ganancia y pérdida y fenómenos de ruido. En la práctica se han utilizado márgenes de enlace que varían de 0 a 6 dB, lo que refleja la gran variación del nivel de certidumbre con el cual es posible modelar los diversos enlaces de comunicación.

Ejemplo de cálculo de un presupuesto de enlace

Como ilustración del cálculo de un presupuesto de enlace, considérese la estación espacial (EE) propuesta como sistema de comunicaciones para atender un sa-

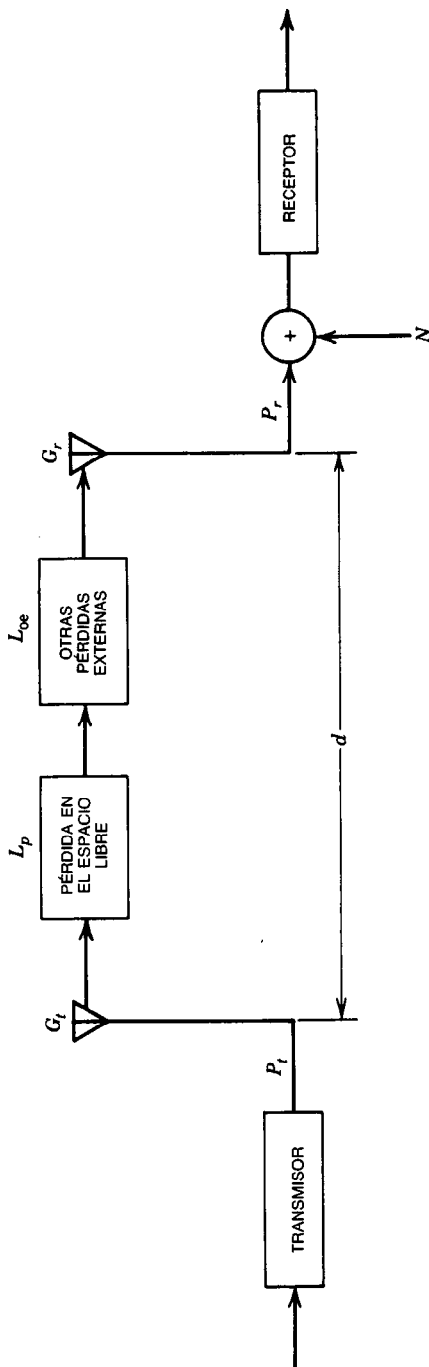


Fig. 48-82. Modelo simplificado de un enlace de comunicaciones.

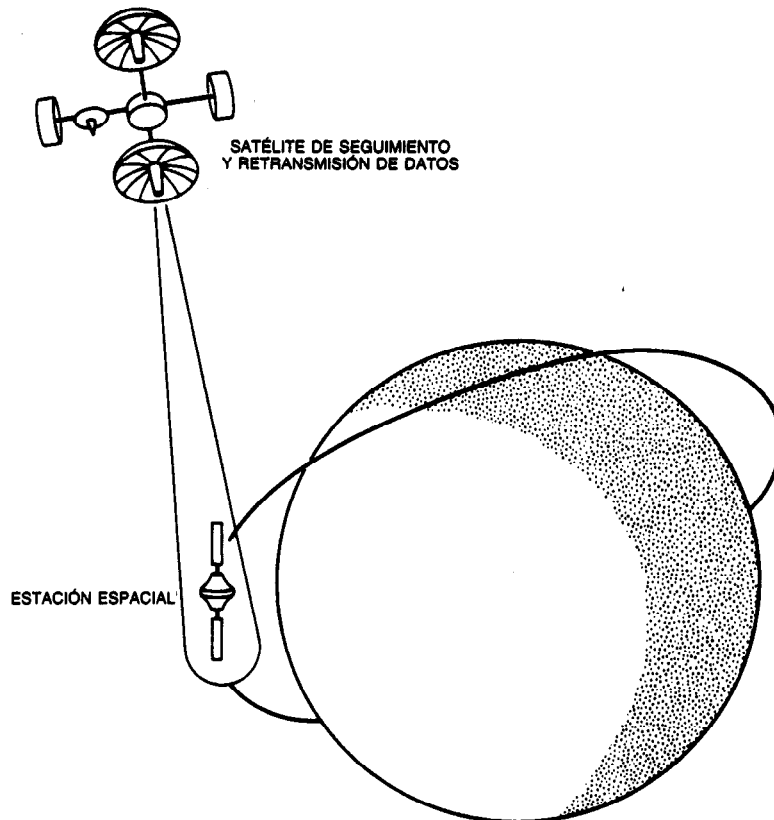


Fig. 48-83. Estación espacial propuesta como sistema de comunicaciones para atender un satélite de seguimiento y retransmisión de datos.

télite de seguimiento y retransmisión de datos (TDRS, de *tracking and data relay satellite*) que se muestra en la figura 48-83. La concepción inicial demanda que el sistema cuente con tres enlaces de comunicación por separado en tres bandas de frecuencia (S, Ku y W). Como ejemplo, el análisis se restringirá a los enlaces de comunicación en la banda Ku.

La estación espacial está planificada para ser puesta en servicio en una órbita terrestre baja, de aproximadamente 270 millas náuticas por encima de la superficie. Entre otras cosas, se comunicará con el TDRS, localizado en una órbita geosíncrona a aproximadamente 22 000 millas náuticas por encima de la superficie del planeta. El TDRS actuará como una estación retransmisora entre la estación espacial y las estaciones terrestres. En la figura 48-84 se ilustra esta función y se separa el sistema de la banda Ku en enlaces ascendente (de la estación terrestre a la espacial) y descendente (de la estación espacial a la terrestre).

Enlace ascendente. En la tabla 48-10 se presenta el presupuesto para el enlace ascendente, en el que se han utilizado las definiciones de los parámetros y las ecuaciones de enlace presentadas previamente. Este ejemplo ilustra el cálculo de un presupuesto de enlace. Existen tres sumas parciales (subrayadas en la tabla).

La estrategia es determinar la primera suma parcial (punto 10), luego la segunda suma parcial (punto 12), y después la suma final (punto 14), en este orden. Pero antes de que sea posible calcular una suma parcial, deben calcularse (en caso necesario) sus diversos elementos, convertirse a decibeles o ambas cosas. Para la primera suma parcial (punto 10), el primer elemento por calcular es la "pérdida de espacio libre" (punto 2). El valor de esta pérdida puede obtenerse a partir de la ecuación 48-81:

$$L_p \text{ (dB)} = 32.5 + 20 \log f + 20 \log d = 206.8 \text{ dB}$$

El segundo punto por calcular es la "temperatura de ruido del sistema receptor de la EE" (punto 6), que se especifica como 440 K, ya que corresponde a un receptor cuyo factor de ruido es de 4 dB:

$$T_e \text{ (dB/K)} = 10 \log(440) = 26.4 \text{ dB/K}$$

El siguiente elemento es el " G/T_e de la EE" (punto 7), que convertido en decibeles es:

$$10 \log (G/T_e) = 10 \log G - 10 \log T_e$$

$$G/T_e \text{ (dB/K)} = G_{\text{dB}} - 26.4 \text{ dB/K}$$

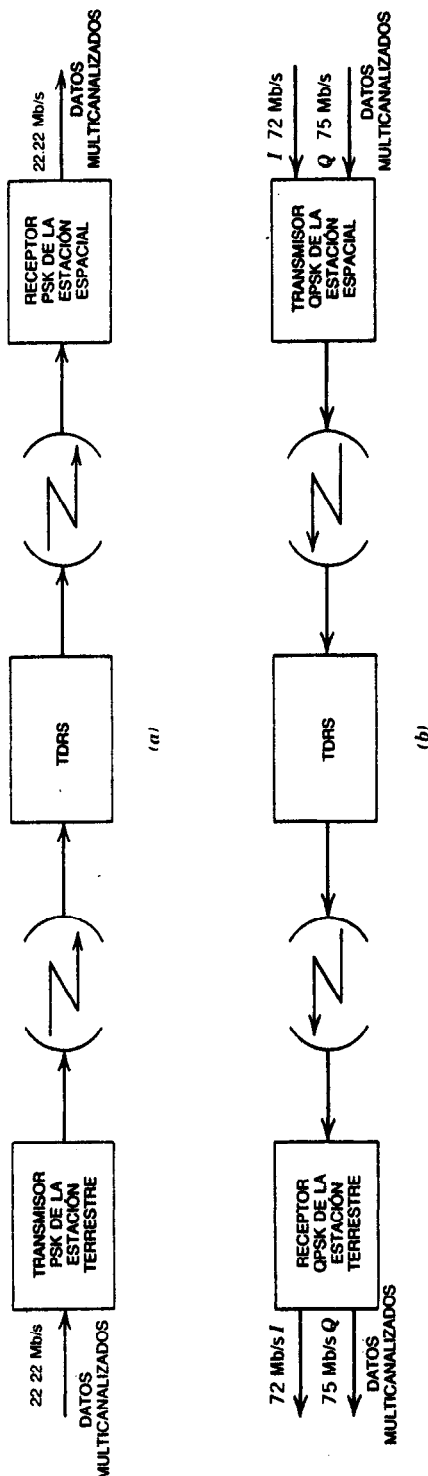


Fig. 48-84. Descomposición del sistema de comunicaciones en la banda Ku, formado por una estación espacial y un satélite de seguimiento y retransmisión de datos (TDRS), en enlaces a) ascendente y b) descendente. PSK = modulación por desplazamiento de fase; QPSK = PSK en cuadratura.

A continuación es necesario evaluar la constante de Boltzmann (punto 8) en decibeles. Se tiene:

$$K = 1.38 \times 10^{-23} \text{ J/K} = -228.6 \text{ dBW/K/Hz}$$

Ahora ya es posible calcular la primera suma parcial, " P_r/N_0 de la EE" (punto 10). Se tiene:

$$\begin{aligned} EE P_r/N_0 \text{ (dBHz)} &= \text{TDRS EIRP (dBW)} - \text{pérdida} \\ &\quad (\text{atenuación}) \text{ en el espacio libre (dB)} - \text{pérdida por direccionamiento} \\ &\quad \text{de la antena (dB)} - \text{pérdida por polarización de la antena (dB)} + EE G/T_e \text{ (dB/K)} - \\ &\quad \text{constante de Boltzmann (dBW/K/Hz)} - \text{pérdida en el circuito receptor de la EE (dB)} = 46.5 - \\ &\quad 206.8 - 0.5 - 0.2 + (G - 26.4) - (-228.6) - 2.5 = 38.7 + G \end{aligned} \quad (48-122)$$

Para el cálculo de la segunda suma parcial se requiere la evaluación de la velocidad de transferencia de los bits en decibeles. Esta velocidad, R , está dada como 22.22 Mb/s en el enlace ascendente, o bien:

$$R \text{ (dBHz)} = 10 \log(22.22 \times 10^6) = 73.5 \text{ dBHz}$$

La segunda suma parcial (en realidad una diferencia), " E_b/N_0 real" (punto 12) está dada entonces por la expresión:

$$\begin{aligned} E_b/N_0 \text{ real (dB)} &= EE P_r/N_0 \text{ (dBHz)} - R \text{ dBHz} \\ &= 38.7 + G - 73.5 = -34.8 + G \text{ (dB)} \end{aligned} \quad (48-123)$$

La suma final (diferencia, en realidad) proporciona el margen de enlace M (punto 14):

$$\begin{aligned} M \text{ (dB)} &= E_b/N_0 \text{ real (dB)} - E_b/N_0 \text{ requerido (dB)} \\ M \text{ (dB)} &= -34.8 + G - 12.3 = -47.1 + G \text{ (dB)} \end{aligned} \quad (48-124)$$

Si se especifica un margen del enlace igual a 3 dB, entonces es posible determinar la incógnita "ganancia de la antena receptora de la EE", G :

$$M = 3 \text{ dB} = -47.1 + G$$

y:

$$G = 3 + 47.1 = 50.1 \text{ dB}$$

Enlace descendente. En la tabla 48-11 se proporciona el cálculo del presupuesto del enlace descendente. Básicamente, es necesario calcular cuatro sumas parciales, que son las expresiones subrayadas de la tabla. La primera suma parcial es la "EIRP de la EE" (punto 4), que está dada por:

Tabla 48-10. Cálculo del presupuesto del enlace ascendente en la banda Ku de la estación espacial propuesta como sistema de comunicación para atender un satélite de seguimiento y retransmisión de datos (TDRS)

Punto	Variable	Suma
1	TDRS EIRP	46.5 dBW
2	Pérdida en el espacio libre L_p ($f = 13.775$ GHz, $d = 38.000$ km)	206.8 dB
3	Pérdida por el direccionamiento de la antena	0.5 dB
4	Pérdida por la polarización de la antena	0.2 dB
5	Ganancia de la antena receptora de la EE, G	G dB
6	Temperatura de ruido del sistema receptor de la EE, T_e (440 K)	26.4 dBK
7	G/T_e de la EE	$G - 26.4$ dBK
8	Constante de Boltzmann, K	$- 228.6$ dBW/K-Hz
9	Pérdida en el circuito receptor de la EE	2.5 dB
10	P_r/N_0 de la EE	$38.7 + G$ dBHz
11	Velocidad de transferencia de los bits, R (22.22 Mb/s)	73.5 dBHz
12	E_b/N_0 verdadero	$- 34.8 + G$ dB
13	E_b/N_0 necesario (para una BER ^a de 10^{-5})	12.3 dB
14	Margen de enlace, M^b	$- 47.1 + G$ dB

^a BER = proporción de errores en los bits.

^b Para $M = 3$ dB, $G = 50.1$ dB.

Tabla 48-11. Cálculo del presupuesto del enlace descendente en la banda Ku de la estación espacial propuesta como sistema de comunicación para atender un satélite de seguimiento y retransmisión de datos (TDRS)

Punto	Variable	Suma
1	Potencia de transmisión de la EE, P_t	P_t dBW
2	Pérdida por el circuito transmisor de la EE	2.5 dB
3	Ganancia de la antena transmisora de la EE, G	50.1 dB
4	EIRP de la EE	$47.6 + P_t$ dBW
5	Pérdida en el espacio libre L_p ($f = 15.0034$ GHz, $d = 38\,000$ km)	207.6 dB
6	Pérdida por el direccionamiento de la antena	0.5 dB
7	Pérdida por la polarización de la antena	0.2 dB
8	G/T_e del TDRS	23.1 dB/K
9	Constante de Boltzmann, K	$- 228.6$ dBW/K-Hz
10	Pérdida por intermodulación del TDRS y otras pérdidas internas	7.0 dB
11	P_r/N_0 del TDRS	$84.0 + P_t$ dBHz
12	Velocidad de transferencia de bits R (75 Mb/s)	78.8 dBHz
13	E_b/N_0 verdadero	$5.2 + P_t$ dB
14	E_b/N_0 necesario (para una BER ^a de 10^{-5})	13.5 dB
15	Margen de enlace, M^b	$- 8.3 + P_t$ dB

^a BER = proporción de errores en los bits.

^b Para $M = 3$ dB, $P_t = 11.3$ dBW (13.5 W).

EE EIRP = potencia de transmisión de la EE, P_t (dBW) - pérdida en el circuito transmisor de la EE + ganancia de la antena transmisora de la EE (dB)

$$= P_t - 2.5 + 50.1 = 47.6 + P_t \text{ dBW} \quad (48-125)$$

donde se establece la hipótesis de que la ganancia de la antena de transmisión y de recepción de la estación espacial son iguales a 50.1 dB, como se determinó en el cálculo del presupuesto del enlace ascendente.

Para el resto de los cálculos del presupuesto del enlace descendente se siguen los mismos pasos que para el caso del enlace ascendente.

Sin embargo, en el cálculo del enlace descendente, el margen del enlace se calcula en términos de la incógnita "potencia de transmisión de la EE", P_t . Al elegir un margen de enlace igual a 3 dB, es posible determinar el valor de P_t :

$$M' = 3 \text{ dB} = -8.3 + P_t$$

$$P_t = 3 + 8.3 = 11.3 \text{ dBW (13.5 W)}$$

REFERENCIAS BIBLIOGRÁFICAS

1. G. Kennedy, *Electronic Communication Systems*, McGraw-Hill, Nueva York.
2. *Reference Data for Radio Engineers*, 5.^a ed., Howard W. Sams, Indianapolis.
3. P. F. Panter, *Communication Systems Design: Line-of-Sight and Tropo-Scatter Systems*, McGraw-Hill, Nueva York.
4. A. W. Rudge, K. Milne, A. D. Olver y P. Knight, eds., *The Handbook of Antenna Design*, vol. 1, Peter Peregrinus, Londres.
5. C. A. Balanis, *Antenna Theory Analysis and Design*, Harper & Row, Nueva York.
6. L. V. Blake, *Antennas*, Wiley, Nueva York.
7. J. J. DeFrance, *Communications Electronics Circuits*, Holt, Rinehart y Winston, Nueva York.
8. L. J. Giacoletto, *Electronics Designers Handbook*, McGraw-Hill, Nueva York.

48.3 COMUNICACIONES ALÁMBRICAS

Matthew J. Quinn, Jr.

Las modernas comunicaciones de datos implican la comunicación de datos entre computadoras y entre elementos de redes de datos (sección 49.3, redes de comunicaciones). Las comunicaciones de datos son esenciales para la operación de redes de procesamiento distribuidas, que se han vuelto cada vez más comunes con la introducción de computadoras más pequeñas y económicas.

Dado que el sistema telefónico nacional estadounidense interconecta aproximadamente 125 millones de teléfonos, su empleo para interconectar computadoras y redes de datos es lógico. Las técnicas y dispo-

sitivos para utilizar el sistema telefónico como red de comunicaciones de datos son el tema de esta sección. En virtud de que anteriormente el sistema telefónico estuvo interconectado completamente por alambres, estas técnicas suelen denominarse comunicaciones alámbricas, aun cuando actualmente en el sistema ya resultan habituales los enlaces por microondas y los satélites.

48.3.1 Definiciones

Un **modem** es un dispositivo utilizado para conectar una fuente o un receptor de datos a una línea telefónica. El nombre "modem" se deriva de los nombres de dos funciones realizadas por el dispositivo, modulación y desmodulación (de *modulation-demodulation*). (En la sección 48.1 puede encontrarse un tratamiento general de estas técnicas.) Los modems se analizan con más detalle posteriormente en esta sección.

Los enlaces de comunicaciones de datos pueden clasificarse de tres modos:

- **Modo simplex.** En éste existe un único enlace en un solo sentido entre el transmisor y el receptor; es decir, la comunicación es solamente unidireccional.
- **Modo dúplex (o dúplex completo).** En él es posible la comunicación simultánea entre los dos extremos del enlace.
- **Modo semidúplex.** Esta operación permite las comunicaciones en ambas direcciones, pero no simultáneamente.

En la figura 48-85 se ilustran los tres modos de operación. Dado que la red telefónica suele ser una red dúplex, en la mayor parte de las comunicaciones de datos se utiliza este modo.

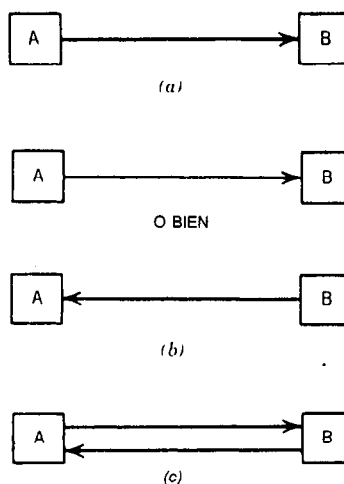


Fig. 48-85. Operación simplex y dúplex de los enlaces de comunicaciones de datos; a) simplex (en un sentido); b) semidúplex (en dos sentidos, no simultáneos); c) dúplex (en dos sentidos, simultáneos).

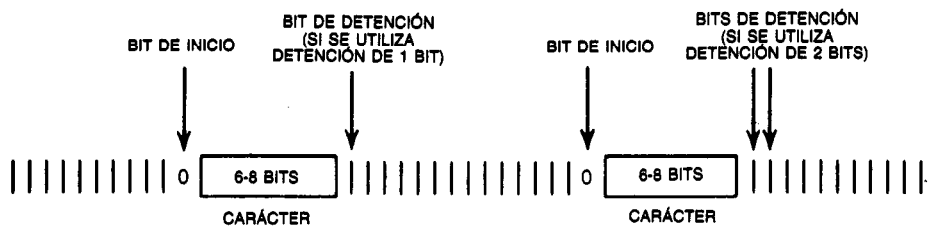


Fig. 48-86. Formato de datos asincrónicos.

Los enlaces pueden subdividirse en sincrónicos y asincrónicos, dependiendo de la forma en que se transmiten y sincronizan los datos en las líneas. En el caso de la transmisión sincrónica, el transmisor y el receptor se sincronizan de modo que bloques de datos de longitud arbitraria puedan enviarse sin interrupción. La transmisión sincrónica se utiliza típicamente en los enlaces de datos de alta velocidad, tales como los enlaces de computadora a computadora. En estos casos el modem transmisor transmite los datos a velocidad constante, y el modem receptor se sincroniza con la corriente de datos de entrada a fin de generar una señal de reloj de datos local. Una vez sincronizado el modem receptor, la transferencia de datos puede continuar durante mucho tiempo. Tales enlaces pueden funcionar de forma muy eficiente (como se define con posterioridad). A menudo se agrega un código para detección de errores, como por ejemplo, un código de redundancia cíclico (CRC, de *cyclic redundancy code*), a fin de verificar que los datos se han transmitido sin errores.

La transmisión asincrónica es común en el enlace entre una sola terminal sin almacenamiento intermedio, por ejemplo, un teletipo que envía caracteres (uno a la vez) con una cantidad arbitraria de tiempo entre cada uno de ellos. Ya que los caracteres no constituyen una corriente continua de datos o siquiera un gran bloque de datos, el modem receptor debe obtener sincronización en cada carácter por separado. En la figura 48-86 se muestra un formato común de datos asincrónicos. Cada carácter empieza con un bit de inicio (cero) y continúa con el código de caracteres de seis, siete u ocho caracteres, dependiendo del código específico que se utilice. La transmisión termina con un patrón de detención de 1, 1 1/2 o 2 bits que por lo general son unos.

Una ventaja de la transmisión asincrónica es que la terminal transmisora puede ser muy sencilla. En la mayor parte de los casos únicamente se requiere un acumulador intermedio (memoria intermedia o "buffer") de un solo carácter. En particular, para la transmisión asincrónica suelen construirse dispositivos terminales de bajo costo.

Codificación

En los enlaces sincrónicos y asincrónicos se utilizan códigos ASCII (American Standard Code for Information Interchange). El código de 7 bits permite 128 combinaciones posibles, como se muestra en las tablas 48-12 y 48-13. Existen 96 caracteres imprimibles y

32 caracteres de control. En equipo IBM suele utilizarse EBCDIC (de *extended binary coded decimal interchange code*, código extendido de intercambio decimal codificado en binario), un código de 8 bits que permite 256 combinaciones posibles, como se muestra en la tabla 48-14. El código de Baudot es un antiguo esquema de codificación para teletipo en el que se utilizan 5 bits, que permiten solamente 32 posibles combinaciones. Uno de los caracteres se utiliza como cambio a mayúsculas, por lo que es posible obtener 64 combinaciones para imprimir 26 caracteres del alfabeto y 10 caracteres numéricos. En la tabla 48-15 se muestra un código de Baudot.

Eficiencia

La eficiencia de un sistema de transmisión de datos puede definirse como la fracción (o porcentaje) de cada sucesión o bloque de transmisión que está constituida realmente por datos. Por ejemplo, un código asincrónico con caracteres de 6 bits, con 1 bit de inicio y 1 bit de detención, tiene eficiencia máxima de:

$$\eta = (6 \text{ bits de datos}) / (6 \text{ bits de datos} + 2 \text{ bits de sincronización}) = 0.75$$

Esta eficiencia máxima se da cuando los caracteres se transmiten en una sucesión continua. Si en vez de 1 bit de detención se utilizan 2 bits, la eficacia se reduce a $6/9 = 0.67$.

La transmisión sincrónica puede hacerse más eficiente que la transmisión asincrónica, ya que los datos pueden enviarse en grandes bloques de caracteres. Son comunes longitudes de bloques de 1 000 bits, con sucesiones de sincronización tan pequeñas como de 24 bits. En este caso, la eficacia es:

$$\eta = 1\,000 / 1\,024 = 0.976$$

Por consiguiente, es posible obtener eficiencias mayores del 99% con transmisión sincrónica. Si se utilizan técnicas para el control de errores como las descritas en la sección 48.5, la eficiencia aparente será menor, pero en tal caso no es el interés principal la eficiencia como se define en esta sección.

48.3.2 Modems para datos

Un **modem para datos** es un dispositivo que acepta datos digitales de una computadora o terminal y los convierte en una forma más adecuada para su transmi-

Tabla 48-12. Conjunto de códigos ASCII de 7 bits

[illegible]

Tabla 48-13. Formatos del código ASCII

De 8 bits			De 7 bits			De paridad par			De paridad impar		
Binario	Hex		Binario	Hex		Binario	Hex		Binario	Hex	
:	10 111 010	BA	:	0 111 010	3A	:	00 111 010	3A	:	10 111 010	BA
;	10 111 011	BB	;	0 111 011	3B	;	10 111 011	BB	;	00 111 011	3B
<	10 111 100	BC	<	0 111 100	3C	<	00 111 100	3C	<	10 111 100	BC
=	10 111 101	BD	=	0 111 101	3D	=	10 111 101	BD	=	00 111 101	3D
>	10 111 110	BE	>	0 111 110	3E	>	10 111 110	BE	>	00 111 110	3E
?	10 111 111	BF	?	0 111 111	3F	?	00 111 111	3F	?	10 111 111	BF
@	11 000 000	CO	@	1 000 000	40	@	11 000 000	CO	@	01 000 000	40
[11 011 011	DB	[1 011 011	5B	[11 011 011	DB	[01 011 011	5B
\	11 011 100	DC	\	1 011 100	BC	\	01 011 100	BC	\	11 011 100	DC
]	11 011 101	DD]	1 011 101	ED]	11 011 101	DD]	01 011 101	ED
^	11 011 110	DE	^	1 011 110	EE	^	11 011 110	DE	^	01 011 110	EE
—	11 011 111	DF	—	1 011 111	SF	—	01 011 111	SF	—	11 011 111	DF
\			\	1 100 000	60	\	01 100 000	60	\	11 100 000	70
{	11 111 011	FB	{	1 111 011	7B	{	01 111 011	7B	{	11 111 011	FB
:	11 111 100	FC	:	1 111 100	7C	:	11 111 100	FC	:	01 111 100	7C
}	11 111 101	FD	}	1 111 101	7D	}	01 111 101	7D	}	11 111 101	FD
~	11 111 110	FE	~	1 111 110	7E	~	01 111 101	7E	~	11 111 110	FE
ACK	10 000 110	B6	ACK	0 000 110	O6	ACK	00 000 110	O6	ACK	10 000 110	86
BEL	10 000 111	B7	BEL	0 000 111	O7	BEL	10 000 111	87	BEL	00 000 111	O7
BS	10 001 000	B8	BS	0 001 000	O8	BS	10 001 000	88	BS	00 001 000	O8
CAN	10 011 000	98	CAN	0 011 000	18	CAN	00 011 000	18	CAN	10 011 000	98
CR	10 001 101	8D	CR	0 001 101	0D	CR	10 001 101	8D	CR	00 001 101	0D
DC1	10 010 001	91	DC1	0 010 001	11	DC1	00 010 001	11	DC1	10 001 001	91
DC2	10 010 010	92	DC2	0 010 010	12	DC2	00 010 010	12	DC2	10 010 010	92
DC3	10 010 011	93	DC3	0 010 011	13	DC3	10 010 011	93	DC3	00 010 011	13
DC4	10 010 100	94	DC4	0 010 100	14	DC4	00 010 100	14	DC4	10 010 100	94
DEL	11 111 111	FF	DEL	1 111 111	7F	DEL	11 111 111	FF	DEL	01 111 111	7F
DLE	10 010 000	90	DLE	0 010 000	10	DLE	10 010 000	90	DLE	00 010 000	10
EM	10 011 001	99	EM	0 011 001	19	EM	10 011 001	99	EM	00 011 001	19
ENQ	10 000 101	85	ENQ	0 000 101	05	ENQ	00 000 101	05	ENQ	10 000 101	85
EOT	10 000 100	84	EOT	0 000 100	04	EOT	10 000 100	84	EOT	00 000 100	04
ESC	10 011 011	9B	ESC	0 011 011	1B	ESC	00 011 011	1B	ESC	10 011 011	9B
ETB	10 010 111	97	ETB	0 010 111	17	ETB	00 010 111	17	ETB	10 010 111	97
ETX	10 000 011	83	ETX	0 000 011	03	ETX	00 000 011	03	ETX	10 000 011	83
FF	10 001 100	8C	FF	0 001 100	0C	FF	00 001 100	0C	FF	10 001 100	8C
FS	10 011 100	9C	FS	0 011 100	1C	FS	10 011 100	9C	FS	00 011 100	1C
GS	10 011 101	9D	GS	0 011 101	1D	GS	00 011 101	1D	GS	10 011 001	99
HT	10 001 001	89	HT	0 001 001	09	HT	00 001 001	09	HT	10 001 001	89
LF	10 001 010	8A	LF	0 001 010	0A	LF	00 001 010	0A	LF	10 001 010	8A
NAK	10 010 101	95	NAK	0 010 101	15	NAK	10 010 101	95	NAK	00 010 101	15
NUL	10 000 000	80	NUL	0 000 000	00	NUL	00 000 000	00	NUL	10 000 000	80
RS	10 011 110	9E	RS	0 011 110	1E	RS	00 011 110	1E	RS	10 011 110	9E
SI	10 001 111	8F	SI	0 001 111	0F	SI	00 001 111	0F	SI	10 001 111	8F

Tabla 48-13. Formatos del código ASCII (continuación)

De 8 bits			De 7 bits			De paridad par			De paridad impar		
Binario		Hex	Binario		Hex	Binario		Hex	Binario		Hex
SO	10 001 110	8E	SO	0 001 110	0E	SO	10 011 110	8E	SO	00 001 110	0E
SOH	10 000 001	81	SOH	0 000 001	01	SOH	10 000 001	81	SOH	00 000 001	01
STX	10 000 010	82	STX	0 000 010	02	STX	10 000 010	82	STX	00 000 010	02
SUB	10 011 010	9A	SUB	0 011 010	1A	SUB	10 011 010	9A	SUB	00 011 010	1A
SYN	10 010 110	96	SYN	0 010 110	16	SYN	10 010 110	96	SYN	00 010 110	16
US	10 011 111	9F	US	0 011 111	1F	US	10 011 111	9F	US	00 111 111	1F
VT	10 001 011	8B	VT	0 001 011	0B	VT	10 001 011	8B	VT	00 001 011	0B

De 8 bits			De 7 bits			De paridad par			De paridad impar		
Binario		Hex	Binario		Hex	Binario		Hex	Binario		Hex
A	11 000 011	C1	A	1 000 001	41	A	01 000 001	41	A	11 000 001	C1
B	11 000 010	C2	B	1 000 010	42	B	01 000 010	42	B	11 000 010	C2
C	11 000 011	C3	C	1 000 011	43	C	11 000 011	C3	C	01 000 011	43
D	11 000 100	C4	D	1 000 100	44	D	01 000 100	44	D	11 000 100	C4
E	11 000 101	C5	E	1 000 101	45	E	11 000 101	C5	E	01 000 101	45
F	11 000 110	C6	F	1 000 110	46	F	11 000 110	C6	F	01 000 110	46
G	11 000 111	C7	G	1 000 111	47	G	01 000 111	47	G	11 000 111	C7
H	11 001 000	C8	H	1 001 000	48	H	01 001 000	48	H	11 001 000	C8
I	11 001 001	C9	I	1 001 001	49	I	11 001 001	C9	I	01 001 001	49
J	11 001 010	CA	J	1 001 010	4A	J	11 001 010	CA	J	01 001 010	4A
K	11 001 011	CB	K	1 001 011	4B	K	01 001 011	4B	K	11 001 011	CB
L	11 001 100	CC	L	1 001 100	4C	L	11 001 100	CC	L	01 001 100	4C
M	11 001 101	CD	M	1 001 101	4D	M	01 001 101	4D	M	11 001 101	CD
N	11 001 110	CE	N	1 001 110	4E	N	01 001 110	4E	N	11 001 110	CE
O	11 001 111	CF	O	1 001 111	4F	O	11 001 111	CF	O	01 001 111	4F
P	11 010 000	D0	P	1 010 000	50	P	01 010 000	50	P	11 010 000	D0
Q	11 010 001	D1	Q	1 010 001	51	Q	11 010 001	D1	Q	01 010 001	51
R	11 010 010	D2	R	1 010 010	52	R	11 010 010	D2	R	01 010 010	52
S	11 010 011	D3	S	1 010 011	53	S	01 010 011	53	S	11 010 011	D3
T	11 010 100	D4	T	1 010 100	54	T	11 010 100	D4	T	01 010 100	54
U	11 010 101	D5	U	1 010 101	55	U	01 010 101	55	U	11 010 101	D5
V	11 010 110	D6	V	1 010 110	56	V	01 010 110	56	V	11 010 110	D6
W	11 010 111	D7	W	1 010 111	57	W	11 010 111	D7	W	01 010 111	57
X	11 011 000	D8	X	1 011 000	58	X	11 011 000	D8	X	11 011 000	58
Y	11 011 001	D9	Y	1 011 001	59	Y	01 011 001	59	Y	11 011 001	D9
Z	11 011 010	DA	Z	1 011 010	5A	Z	01 011 010	5A	Z	11 011 010	DA
a	11 100 001	E1	a	1 100 001	61	a	11 100 001	E1	a	01 100 001	61
b	11 100 010	E2	b	1 100 010	62	b	11 100 010	E2	b	01 100 010	62
c	11 100 011	E3	c	1 100 011	63	c	01 100 011	E3	c	11 100 011	E3
d	11 100 100	E4	d	1 100 100	64	d	11 100 100	E4	d	01 100 100	64
e	11 100 101	E5	e	1 100 101	65	e	01 100 101	65	e	11 100 101	E5
f	11 100 110	E6	f	1 100 110	66	f	01 100 110	66	f	11 100 110	E6
g	11 100 111	E7	g	1 100 111	67	g	11 100 111	E7	g	01 100 111	67
h	11 101 000	E8	h	1 101 000	68	h	11 101 000	E8	h	01 101 000	68
i	11 101 001	E9	i	1 101 001	69	i	01 101 001	69	i	11 101 001	E9
j	11 101 010	EA	j	1 101 010	6A	j	01 101 010	6A	j	11 101 010	EA
k	11 101 011	EB	k	1 101 011	6B	k	11 101 011	EB	k	01 101 011	6B
l	11 101 100	EC	l	1 101 100	6C	l	01 101 100	6C	l	11 101 100	EC

Tabla 48-13. Formatos del código ASCII (continuación)

De 8 bits			De 7 bits			De paridad par			De paridad impar		
	Binario	Hex		Binario	Hex		Binario	Hex		Binario	Hex
m	11 101 101	ED	m	1 101 101	6D	m	11 101 101	ED	m	01 101 101	6D
n	11 101 110	EE	n	1 101 110	6E	n	11 101 110	EE	n	01 101 110	6E
o	11 101 111	EF	o	1 101 111	6F	o	01 101 111	6F	o	11 101 111	EF
p	11 110 000	F0	p	1 110 000	70	p	11 110 000	F0	p	01 110 000	70
q	11 110 001	F1	q	1 110 001	71	q	01 110 001	71	q	11 110 001	F1
r	11 110 010	F2	r	1 110 010	72	r	01 110 010	72	r	11 110 010	F2
s	11 110 011	F3	s	1 110 011	73	s	11 110 011	F3	s	01 110 011	73
t	11 110 100	F4	t	1 110 100	74	t	01 110 100	74	t	11 110 100	F4
u	11 110 101	F5	u	1 110 101	75	u	11 110 101	F5	u	01 110 101	75
v	11 110 110	F6	v	1 110 110	76	v	11 110 110	F6	v	01 110 110	76
w	11 110 111	F7	w	1 110 111	77	w	01 110 111	77	w	11 110 111	F7
x	11 111 000	F8	x	1 111 000	78	x	01 111 000	78	x	11 111 000	F8
y	11 111 001	F9	y	1 111 001	79	y	11 111 001	F9	y	01 111 001	79
z	11 111 010	FA	z	1 111 010	7A	z	11 111 010	FA	z	01 111 010	7A
0	10 110 000	B0	0	0 110 000	30	0	00 110 000	30	0	10 110 000	B0
1	10 110 001	B1	1	0 110 001	31	1	10 110 001	B1	1	00 110 001	31
2	10 110 010	B2	2	0 110 010	32	2	10 110 010	B2	2	00 110 010	32
3	10 110 011	B3	3	0 110 011	33	3	00 110 011	33	3	10 110 011	B3
4	10 110 100	B4	4	0 110 100	34	4	10 110 100	B4	4	00 110 100	34
5	10 110 101	B5	5	0 110 101	35	5	00 110 101	35	5	10 110 101	B5
6	10 110 110	B6	6	0 110 110	36	6	00 110 110	36	6	10 110 110	B6
7	10 110 111	B7	7	0 110 111	37	7	10 110 111	B7	7	00 110 111	37
8	10 111 000	B8	8	0 111 000	38	8	10 111 000	B8	8	00 111 000	38
9	10 111 001	B9	9	0 111 001	39	9	00 111 001	39	9	10 111 001	B9
SP	10 100 000	A0	SP	0 100 000	20	SP	10 100 000	A0	SP	00 100 000	20
!	10 100 001	A1	!	0 100 001	21	!	00 100 001	21	!	10 100 001	A1
~	10 100 010	A2	~	0 100 010	22	~	00 100 010	22	~	10 100 010	A2
#	10 100 011	A3	#	0 100 011	23	#	10 100 011	A3	#	00 100 011	23
\$	10 100 100	A4	\$	0 100 100	24	\$	00 100 100	24	\$	10 100 100	A4
%	10 100 101	A5	%	0 100 101	25	%	10 100 101	A5	%	00 100 101	25
&	10 100 110	A6	&	0 100 110	26	&	10 100 110	A6	&	00 100 110	26
'	10 100 111	A7	'	0 100 111	27	'	00 100 111	27	'	10 100 111	A7
(10 101 000	A8	(0 101 000	28	(00 101 000	28	(10 101 000	A8
)	10 101 001	A9)	0 101 001	29)	10 101 001	A9)	00 101 001	29
*	10 101 010	AA	*	0 101 010	2A	*	10 101 010	AA	*	00 101 010	2A
+	10 101 011	AB	+	0 101 011	2B	+	00 101 011	2B	+	10 101 011	AB
,	10 101 100	AC	,	0 101 100	2C	,	10 101 100	AC	,	00 101 100	2C
-	10 101 101	AD	-	0 101 101	2D	-	00 101 101	2D	-	10 101 101	AD
.	10 101 110	AE	.	0 101 110	2E	.	00 101 110	2E	.	10 101 110	AE
/	10 101 111	AF	/	0 101 111	2F	/	10 101 111	AF	/	00 101 111	2F

Tabla 48-14. Representación hexadecimal de códigos alfanuméricos. (Comunicaciones sincrónicas binarias de IBM)

Asignación de caracteres ASCII				Asignación de caracteres EBCDIC ^a (definidos para la IBM 3270 ^b)				Asignación de caracteres transcódigo de 6 bits	
Carácter	Hex	Carácter	Hex	Carácter	Hex	Carácter	Hex	Carácter	Hex
A	41	"	22	A	C1	*	5C	A	01
B	42	#	23	B	C2	%	6C	B	02
C	43	\$	24	C	C3	@	7C	C	03
D	44	%	25	D	C4	(4D	D	04
E	45	&	26	E	C5)	5D	E	05
F	46	'	27	F	C6	—	6D	F	06
G	47	(28	G	C7	.	7D	G	07
H	48)	29	H	C8	+	4E	H	08
I	49	*	2A	I	C9	;	5E	I	09
J	4A	+	2B	J	D1	>	6E	J	11
K	4B	,	2C	K	D2	=	7E	K	12
L	4C	-	2D	L	D3		4F	L	13
M	4D	.	2E	M	D4	—	5F	M	14
N	4E	/	2F	N	D5	?	6F	N	15
O	4F	:	3A	O	D6	"	7F	O	16
P	50	;	3B	P	D7	SBA	11	P	17
Q	51	<	3C	Q	D8	EUA	12	Q	18
R	52	=	3D	R	D9	IC	13	R	19
S	53	>	3E	S	E2	RA	3C	S	22
T	54	?	3F	T	E3	DLE	10	T	23
U	55	@	40	U	E4	EM	19	U	24
V	56	[5B	V	E5	ENQ	2D	V	25
W	57	\	5C	W	E6	ETB	26	W	26
X	58]	5D	X	E7	EOT	37	X	27
Y	59	^	5E	Y	EB	ESC	27	Y	28
Z	5A	—	5F	Z	E9	ETX	03	Z	29
a	61	\	60	a	81	FF	0C	0	30
b	62	{	7B	b	82	PT	05	1	31
c	63		7C	c	83	DUP	IC	2	32
d	64	}	7D	d	84	SF	1D	3	33
e	65	~	7E	e	85	FM	1E	4	34
f	66	BEL	07	f	86	ITB	1F	5	35
g	67	BS	08	g	87	NAK	3D	6	36
h	68	CAN	18	h	88	NL	15	7	37
i	69	CR	0D	i	89	NUL	00	8	38
j	6A	DC1	11	j	91	ESC	27	9	39
k	6B	DC2	12	k	92	SOH	01	Espacio	1A
l	6C	DC3	13	l	93	Espacio	40		0B

Tabla 48-14. Representación hexadecimal de códigos alfanuméricos. (Comunicaciones sincrónicas binarias de IBM) (continuación)

Asignación de caracteres ASCII				Asignación de caracteres EBCDIC ^a (definidos para la IBM 3270 ^b)				Asignación de caracteres transcódigo de 6 bits	
Carácter	Hex	Carácter	Hex	Carácter	Hex	Carácter	Hex	Carácter	Hex
m	6D	DC4	14	m	94	STX	02	.	2B
n	6E	DEL	7F	n	95	SUB	3F	\$	1B
o	6F	DLE	10	o	96	SYN	32	#	3B
p	70	EM	19	p	97			,	0C
q	71	ENQ	05	q	98			+	1C
r	72	EOT	04	r	99			/	21
s	73	ESC	1B	s	A2			%	2C
t	74	ETB	17	t	A3			@	3C
u	75	ETX	03	u	A4			-	20
v	76	FF	0C	v	A5			&	10
w	77	FS	1C	w	A6			BEL	0D
x	78	GS	1D	x	A7			DEL	3F
y	79	HT	09	y	A8			DLE	1F
z	7A	LF	0A	z	A9			EM	3E
0	30	NAK	15	0	F0			ENQ	2D
1	31	NUL	00	1	F1			EOT	1E
2	32	RS	1E	2	F2			ESC	2A
3	33	S1	0F	3	F3			ETB	0F
4	34	SO	0E	4	F4			ETX	2E
5	35	SOH	01	5	F5			HT	2F
6	36	STX	02	6	F6			NAK	3D
7	37	SUB	1A	7	F7			SOH	00
8	38	SYN	16	8	F8			STX	0A
9	39	US	1F	9	F9			SUB	0E
Espacio	20	VT	0B	&	50			SYN	3A
!				/	61			US	1D
				\$	5B				
				¢	4A				
				!	5A				
					6A				
				:	7A				
				#	7B				
				,	6B				
				.	4B				
				<	4C				

^a EBCDIC = código extendido de intercambio decimal codificado en binario.

^b Otras terminales tendrán funciones definidas que no aparecen aquí, por ejemplo, el hexadecimal 04 para la IBM 2780 es PF. De forma parecida, otras terminales pueden tener nombres diferentes para funciones definidas aquí, por ejemplo, el hexadecimal 05 para la IBM 2780 es HT.

Tabla 48-15. Formato del código Baudot

Carácter		Posición de la marca				
Letra minúscula	Letra mayúscula					
A	—					
B	?					
C	:					
D	\$					
E	3					
F	!					
G	&					
H	#					
I	8					
J	'					
K	(
L)					
M	.					
N	,					
O	9					
P	0					
Q	1					
R	4					
S	Bell					
T	5					
U	7					
V	;					
W	2					
X	/					
Y	6					
Z	"					
Letras (cambio a letras minúsculas)						
Números (cambio a letras mayúsculas)						
Espacio						
Regreso de carro						
Avance de renglón						
Espacio en blanco						

sión por la red telefónica. En el extremo receptor otro modem restituye los datos a su forma original. Los modems suelen denominarse equipos (receptor/transmisor) de datos y Dataphones en las compañías telefónicas.

Antes del 1.º de enero de 1969, en la red telefónica pública de Estados Unidos solamente era posible utilizar modems o Dataphones de la compañía telefónica. En esa época entró en vigor la decisión Carterfone, con lo que se permitió la conexión de accesorios externos en la red telefónica. A pesar de ello, era necesario conectar un DAA (circuito de acceso de datos, de *data access arrangement*) suministrado por la compañía Bell, a fin de proteger el sistema telefónico.

A partir de entonces, la FCC ha certificado muchos otros modems, lo que permite su conexión al sistema telefónico sin DAA.

48.3.3 Técnicas de modulación para modems

Existen tres formas de modular una señal portadora: haciendo variar su amplitud, su fase o su frecuencia. Estas formas se refieren a la AM, PM o FM. Para datos binarios estos tipos de modulación se denominan ASK, PSK y FSK (manipulación por desplazamiento de amplitud, manipulación por desplazamiento de fase —defasamiento— y manipulación por desplazamiento de frecuencia, respectivamente).

- **ASK.** Suele utilizarse para enlaces por fibras ópticas.
- **FSK.** Normalmente se utiliza en enlaces asincrónicos, ya que es posible desmodularla asincrónicamente. Es el sistema más fácil de construir y el más económico. Por consiguiente, es el sistema ideal para operación a baja velocidad. Es un sistema bastante sólido, lo que es importante cuando se utiliza en redes de computadoras operacionales. La desventaja de la FSK es el ancho de banda relativamente grande que se emplea.
- **PSK.** Es el método más eficiente para transmitir datos binarios en presencia de ruido. La principal desventaja de la PSK es la complejidad del transmisor y del receptor, en comparación con lo que ocurre en el caso de la FSK. La señal PSK debe desmodularse sincrónicamente, por lo que es una candidata ideal para el modem sincrónico. Los requisitos de ancho de banda son excelentes si se emplea QSPK (PSK en cuadratura). Algunas veces se utiliza DPSK (PSK diferencial) a fin de simplificar el receptor.
- **QAM.** La modulación de amplitud en cuadratura es un método bastante eficiente para obtener el máximo rendimiento de un ancho de banda limitado. Es una combinación de manipulación por desplazamiento de fase y de amplitud.

En el transmisor PSK se utilizan dos fases para enviar un uno y un cero, como se muestra en la figura 48-87a. A fin de mejorar la eficiencia, los transmisores PSK pueden emplear cuatro fases para enviar los cuatro números 00, 01, 10, 11, como se muestra en la figura 48-87b. Las señales se codifican utilizando el código Gray.

La eficiencia se incrementa aún más mediante la transmisión de 8 fases para enviar los ocho números 000, 001, 010, 011, 100, 101, 110, 111. La QAM suele utilizarse para tales transmisiones, como se observa en la figura 48-87c. En la figura 48-87d ocho fases y dos amplitudes caracterizan una QAM de ocho fases, que tiene 16 asignaciones de señales, cada una correspondiente a un número de 4 bits. La asignación de la fase y la amplitud de la señal conforme al código Gray asegura que solamente existe un bit de diferencia entre fases adyacentes. Por consiguiente, si se detecta un error de fase, la probabilidad es que sólo un bit esté equivocado.

48.3.4 Velocidades de transmisión de datos

Las velocidades de transmisión de datos de los modems empiezan a partir de 50 b/s (bits por segundo), y se elevan hasta el intervalo de los megabits por segundo (Mb/s). En esta sección sólo se analizarán los modems comunes de la banda del habla (o de calidad telefónica), susceptibles de conectarse a la red telefónica. Por lo general estos modems están limitados a 9 600 b/s y emplean aproximadamente 3 kHz de ancho de banda. Con una razón de señal sobre ruido (SNR) aproximada de 30 dB y ancho de banda de unos 3 kHz,

las velocidades de transmisión de datos en la vecindad de 30 kb/s (kilobits por segundo) son teóricamente posibles (sección 47.3, Teoría de la información). Sin embargo, el canal telefónico no es un ejemplo del conocido canal de ruido gaussiano blanco aditivo (AWGN, de *additive white Gaussian noise*). El canal telefónico está caracterizado como un canal de error irruptivo o explosivo, ya que la red es afectada por efectos transitorios de la conmutación en el canal del habla, es

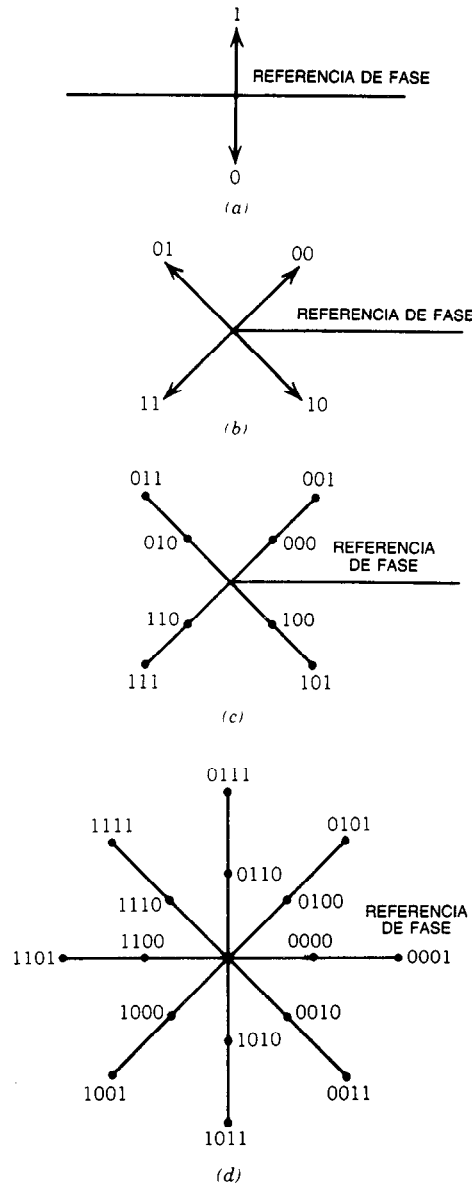


Fig. 48-87. Esquemas de modulación de fase en los modems para datos: a) manipulación por desplazamiento de fase binaria (BPSK); b) PSK en cuadratura (QPSK); c) modulación de amplitud en cuadratura (QAM) de cuatro fases; d) QAM de ocho fases.

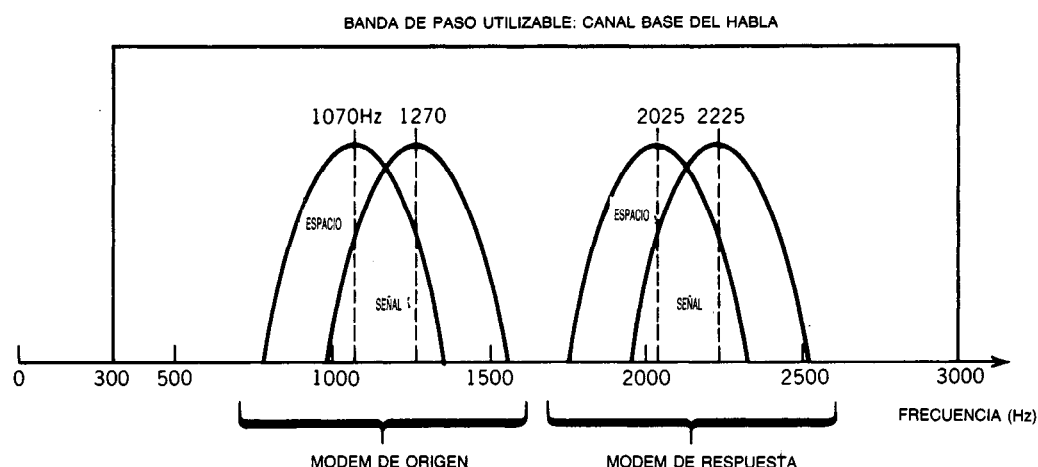


Fig. 48-88. Espectro de frecuencias para un modem dúplex.

objeto de una gran cantidad de interferencia, y es vulnerable a perturbaciones atmosféricas debido a los enlaces de microondas en el sistema telefónico. Como resultado, la máxima velocidad de transmisión de datos de uso común para los modems es de 9 600 b/s.

Después de unos 30 años de demanda por los modems y de más de 12 años de competencia en el campo de los modems de la banda del habla, ha perdurado una familia de modems compatibles. Las diversas velocidades permiten al usuario una elección razonable de tipos y costos, a fin de que la eficiencia de costo determine cuál modem utilizar en una aplicación específica. Existen modems con velocidades de transmisión de 300, 1 200, 1 800, 2 400, 4 800 y 9 600 b/s.

En el caso de los modems para datos de baja velocidad, la velocidad de transmisión de datos (b/s) suele ser la misma que la velocidad de manipulación (símbolos por segundo, s/s). Esta razón de 1 bit por símbolo puede utilizarse hasta 2 400 b/s, aunque el ancho de banda del canal telefónico de habla básico de 2 700 Hz es menor que el espectro de potencia transmitida a 2 400 b/s. Se ha encontrado que la cantidad de distorsión en los símbolos de datos recibidos en el receptor es aceptable a 2 400 b/s (Davey). Para la comunicación de datos a baja velocidad, en el modem dúplex típico

se utiliza la FSK, con las frecuencias de transmisión y recepción dispuestas como en la figura 48-88.

Para velocidades de transmisión de datos mayores de 2 400 b/s, se requieren técnicas de modulación más complicadas (fig. 48-87). En estos esquemas suelen utilizarse técnicas de modulación en fase coherente. Si se utiliza la PSK de cuatro fases, existen dos bits por símbolo, y es posible una velocidad de transmisión de datos de 4 800 b/s. De forma semejante, si se emplea QAM de ocho fases (fig. 48-87d), existen 16 estados de señales, o 4 bits por símbolo. Con este formato de modulación, es posible transmitir 9 600 b/s a través de un canal telefónico de habla. En la tabla 48-16 se resumen las velocidades alcanzables de transmisión de datos para los diferentes formatos de modulación. Para las más altas velocidades de transmisión, se requieren líneas privadas no conmutadas.

48.3.5 Tipos de servicio telefónico

Para la transmisión de datos existen varios tipos de servicio telefónico. Uno es la línea privada. La interconexión para esta línea es muy sencilla, ya que es una conexión de tiempo completo. No es necesario realizar muchas de las funciones normales de una llamada co-

Tabla 48-16. Características comunes de la transmisión de datos

Tipo de modulación	Bits/símbolo	Bits/s en el canal del habla	Modo	Requisito de línea
FSK binaria ^a	1	1200	ASYN	DDD ^b
PSK binaria	1	2400	SYN	DDD/PL
PSK de cuatro fases	2	4800	SYN	DDD/PL
QAM de cuatro fases	3	7200	SYN	PL
QAM de ocho fases	4	9600	SYN	PL

^a FSK = manipulación por desplazamiento de frecuencia; PSK = manipulación por desfaseamiento; QAM = modulación de amplitud en cuadratura.

^b DDD = marcado directo a distancia; PL = línea privada.

riente, y tampoco es necesario estar conectado a la línea, esperarla y luego colgar el auricular cuando se ha terminado la operación. Todo lo que se requiere son tres señales: listo para enviar, solicitud de transmisión y detección de portadora.

El tipo más sencillo de circuito es la línea privada, aunque la conexión de tiempo completo puede resultar costosa. Si el circuito se utiliza todo el tiempo, el costo justifica el lujo de tener una conexión permanente. Sin embargo, si el circuito se utiliza sólo algunas horas del día es más económico un modem para marcar (llamar) o DDD (de *direct distance dialing*, marcado directo a distancia). No se requiere una cantidad específica de uso para justificar la decisión entre contar con conexión de tiempo completo o marcar; el usuario debe evaluar los cargos por el servicio como función del costo de la conexión de tiempo completo.

Como un modem DDD, el usuario debe considerar aspectos como la contestación automática, la desconexión automática y el problema de las designaciones

de expedición respuesta. En ciertos modems dúplex, el modem expeditor (de origen) transmite una señal (uno) a 1 270 Hz y un espacio (cero) a 1 070 Hz, y recibe una señal a 2 225 Hz y un espacio a 2 025 Hz. El modem de respuesta debe manipular, evidentemente, las frecuencias contrarias. Algunas veces es posible utilizar un modem indistintamente, como el de expedición o el de respuesta. Esta capacidad debe estar incorporada al modem.

48.3.6 Acondicionamiento de líneas

Otra consideración importante es la necesidad de acondicionamiento. A medida que se han mejorado las técnicas de modulación y se han desarrollado las de desmodulación ha sido menos necesario el acondicionamiento de líneas, debido a que los modems receptores a menudo cuentan un equipo de ecualización integrado en el sistema. Con él se pretende corregir las distorsiones y no linealidades en el canal.

Tabla 48-17. Acondicionamiento de la línea telefónica para la transmisión de datos

Acondicionamiento del canal ^a	Distorsión por atenuación (respuesta de frecuencia) con respecto a 1 004 Hz		Distorsión por retardo de la envolvente	
	Intervalo de frecuencias (Hz) ^b	Variación (dB) ^d	Intervalo de frecuencias (Hz)	Variación (μs)
Básico	500-2500 300-3000	- 2 a + 8 - 3 a + 12	800-2600	1750
C1	1000-2400 ^c 300-2700 ^c 300-3000	- 1 a + 3 - 2 a + 6 - 3 a + 12	1000-2400 ^c 800-2600	1000 1750
C2	500-2800 ^c 300-3000 ^c	- 1 a + 3 - 2 a + 6	1000-2600 ^c 600-2600 ^c 500-2800 ^c	500 1500 3000
C3 (línea de acceso)	500-2800 ^c 300-3000	- 0.5 a + 1.5 - 0.8 a + 3	1000-2600 ^c 600-2600 ^c 500-2800 ^c	110 300 650
C3 (troncal)	500-2800 ^c 300-3000 ^c	- 0.5 a + 1 - 0.8 a + 2	1000-2600 ^c 600-2600 ^c 500-2800 ^c	80 260 500
C4	500-3000 ^c 300-3200 ^c	- 2 a + 3 - 2 a + 6	1000-2600 ^c 800-2800 ^c 600-3000 ^c 500-3000 ^c	300 500 1500 3000
C5	500-2800 ^c 300-3000 ^c	- 0.5 a + 1.5 - 1 a + 3	1000-2600 ^c 600-2600 ^c 500-2800 ^c	100 300 600

^a El acondicionamiento C se aplica sólo a las características de atenuación y de retardo de la envolvente.

^b Las frecuencias medidas estarán 4 Hz por encima de las que se muestran. Por ejemplo, el canal básico tendrá de - 2 a + 8 dB de pérdida entre 504 y 2 504 Hz.

^c Valores calculados.

^d (+) significa pérdida y (-) significa ganancia con respecto a 1 004 Hz.

Para comprender el acondicionamiento es importante conocer los diferentes parámetros de la red telefónica, que es un sistema enorme y, como tal, bastante difícil de mantener en cuanto a uniformidad. Por consiguiente, la compañía telefónica estadounidense ha publicado las especificaciones de un circuito normal y pretende mantener todos sus circuitos dentro de tales especificaciones.

Las señales de datos suelen estar diseñadas con menos redundancia que las del habla, y es necesario detectarlas con medios electrónicos en vez de con el oído y el cerebro. Por tanto, defectos sencillos en el circuito telefónico que no afectan gravemente la recepción del habla en el receptor pueden ser dañinos para la recepción de datos. Los circuitos telefónicos se han optimizado para la voz y el habla, y muchos de los filtros, que tienen cortes muy pronunciados a fin de asegurar un mínimo de interferencia entre canales adyacentes, afectan negativamente a la transmisión de datos debido a que sus características de fase son muy deficientes. Antes de la gran demanda de los circuitos de datos, las compañías telefónicas no consideraban un problema las características de fase deficientes, debido a que el oído humano es relativamente insensible a la distorsión de fase. De este modo, al proporcionar una respuesta de amplitud plana en la banda de frecuencias, la respuesta de fase no se controlaba con cuidado.

Durante la transmisión de un pulso por un circuito de comunicaciones, el efecto de la distorsión de fase (distorsión de retraso) es que las muchas componentes de la frecuencia del pulso experimentan diferentes cantidades de retraso a través del circuito. En el receptor las componentes no parecen constituir un pulso bien definido. El resultado es un pulso distorsionado, especialmente en los flancos ascendente y descendente. El pulso, que tenía una duración bien definida cuando fue transmitido, se difumina en el tiempo. Debido a este efecto, los pulsos tienden a interferirse mutuamente en el receptor. Este fenómeno se denomina interferencia entre símbolos y puede ser el efecto limitante del rendimiento en un sistema seriamente limitado en banda.

El acondicionamiento de líneas, mediante dispositivos de compensación denominados ecualizadores, se utiliza a fin de reducir los efectos de la distorsión en fase y amplitud en las líneas telefónicas para la transmisión de datos. En la tabla 48-17 se resume una selección estándar de circuitos acondicionados. En ella se muestra para efectos de comparación la línea privada básica de calidad telefónica (tipo 3002).

48.3.7 Conclusiones

Los sistemas implementados mediante el empleo de modems deben tener ciertos objetivos de funcionamiento. Estos objetivos se resumen en la tabla 48-18. Primero, la elección más importante es la velocidad de transmisión de datos, ya que el interés es minimizar el retraso en la mayor parte de los casos. Las velocidades más altas de transmisión presentan retrasos más breves, y por lo general el requisito se expresa en algún tiempo medio de respuesta. En la mayor parte de los sistemas de tiempo real, un tiempo razonable de res-

puesta es de cinco segundos aproximadamente para mantener una conversación con otra persona. También el rendimiento es importante, y la medición típica es un determinado nivel de tráfico pico.

Segundo, la disponibilidad del servicio reviste igual importancia, y suele especificarse en los requisitos. Se define como:

$$\text{Disponibilidad} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

donde MTBF es el tiempo medio entre fallas (de *mean time between failures*) y MTTR es el tiempo medio de restitución (de *mean time to restore*). Por ejemplo, un sistema con MTBF de 1 000 horas y MTTR de 10 horas tendrá disponibilidad de:

$$\text{Disponibilidad} = \frac{1000}{1000 + 10} = 99\%$$

Por otra parte, un sistema con MTBF de 1 000 horas y MTTR de 100 horas tendrá disponibilidad de solamente:

$$\text{Disponibilidad} = \frac{1000}{1000 + 100} = 91\%$$

La integridad de los datos es un factor importante; se mide en BER (de *bit error rate*, proporción de errores en los bits). Un canal de modem normal debe tener un BER de no más de 10^{-6} , lo que significa que un bit en un millón es erróneo. Con una BER promedio de 10^{-6} la proporción de pérdida del mensaje será baja, dependiendo del tamaño del mensaje y de la naturaleza irruptiva o explosiva de los bits erróneos. Ya que es posible que muchos bits erróneos se encuentren acumulados en un mensaje, la proporción de pérdida del mensaje suele ser menor en los canales telefónicos que en los canales de ruido gaussiano blanco aditivo con errores aleatorios.

Tabla 48-18. Objetivos de rendimiento de sistemas de comunicaciones alámbricas en los que se utilizan modems

Componente	Parámetro	Medida típica
Velocidad	Retardo	Tiempo de respuesta promedio
	Rendimiento	Nivel de tráfico pico
Servicio	Disponibilidad	Porcentaje de tiempo de inactividad
	Integridad de los datos	Proporción de errores en los bits
	Integridad del mensaje	Proporción de pérdida del mensaje
	Seguridad	Proporción de entrega equivocada del mensaje

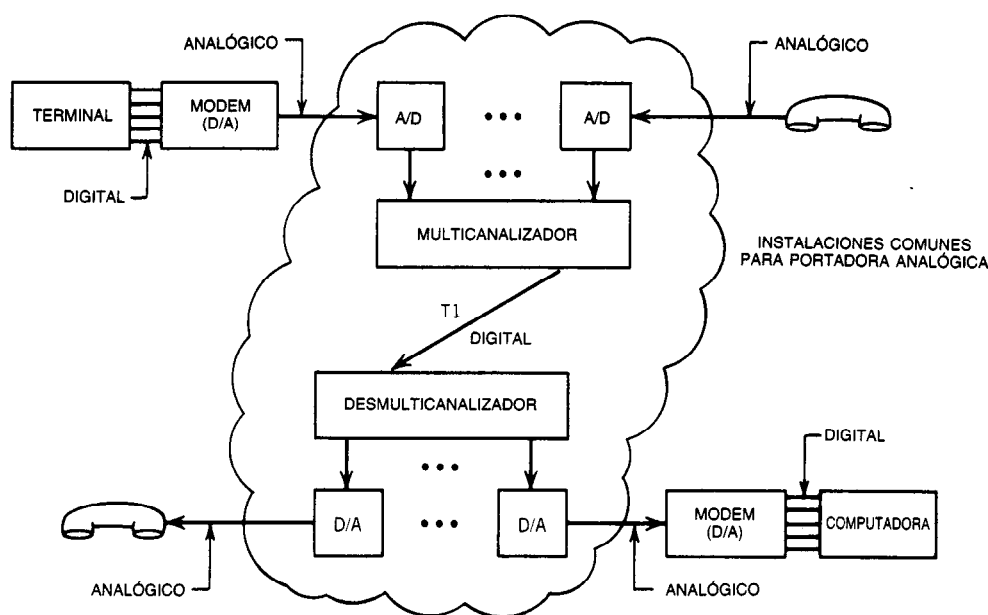


Fig. 48-89. Transmisión digital en la red telefónica.

Por último, la seguridad de los datos es una medida importante para asegurar que los mensajes no serán entregados a destinos equivocados. Esto es particularmente importante en una red de comunicaciones.

Una cantidad significativa de investigación y desarrollo ha sido dedicada al diseño, desarrollo y perfeccionamiento de modems y dispositivos afines para hacer posible la comunicación de datos digitales por un sistema esencialmente analógico. Asimismo, se han dedicado cantidades igualmente importantes de tiempo al desarrollo de técnicas, dispositivos y sistemas para sustituir grandes porciones del sistema telefónico analógico por sistemas digitales. En la figura 48-89 se muestra un diagrama simplificado de comunicaciones digitales multicanalizadas por medio de un sistema portador T1.

Actualmente ya es común que los datos digitales de un usuario se conviertan a una portadora analógica modulada digitalmente, que a su vez se convierte a forma digital para su transmisión. Un receptor digital reconvierte la señal en una portadora modulada, que después es desmodulada por un modem receptor a fin de restituir los datos originales. Si bien esto puede considerarse una forma tortuosa de comunicación de datos, permite que los usuarios de datos y del canal del habla utilicen simultáneamente los amplios recursos del sistema telefónico nacional.

En resumen, se ha demostrado que el empleo de la red telefónica para la comunicación de datos es extraordinariamente efectivo. Redes de computadoras, sistemas de telemetría remota y bases de datos para usuarios de computadoras son sólo tres ejemplos de prácticas que normalmente serían en extremo costosas, si no imposibles, sin el amplio uso de las comunicaciones "alámbricas".

LIBROS Y REVISTAS

- Bell Telephone Labs, *Engineering and Operations in the Bell System*, Indianapolis.
- W. P. Davenport, *Modern Data Communication*, Hayden, Rochelle Park, NJ.
- J. R. Davey, "MODEMS", *Proc. IEEE* 60(11): 1284-1292.
- Lenkhurt Electric Co., *The Lenkhurt Demodulator*, San Carlos, CA.
- J. Martin, *Telecommunications and the Computer*, Prentice-Hall, Englewood Cliffs, NJ.
- J. Martin, *Communications Satellite Systems*, Prentice-Hall, Englewood Cliffs, NJ.
- M. Schwartz, *Computer-Communication Network Design and Analysis*, Prentice-Hall, Englewood Cliffs, NJ.
- W. Sinnema, *Digital, Analog and Data Communication*, Reston, Reston, VA.

48.4 SISTEMAS DE COMUNICACIÓN POR FIBRAS ÓPTICAS

K. K. Chow

48.4.1 Primeros desarrollos de las fibras ópticas

Con toda certeza, la propagación de la luz confinada en un medio transparente es conocida desde hace más de un siglo. Por ejemplo, en 1870, Tyndal demostró ante la Royal Society la propagación de la luz en un chorro de agua¹. Sin embargo, el trabajo formal sobre el uso práctico de ondas de luz guiadas comenzó a principios de la década de 1950. En esa época los centros

de mayor actividad eran Holanda (van Heel) e Inglaterra (Hopkins, Kapany y Hirschowitz), donde se estudiaba la propagación de la luz en fibras sin recubrir, recubiertas de plástico y recubiertas de vidrio para la transmisión de imágenes². De hecho, fueron Kapany y colaboradores³ quienes acuñaron el término "óptica de fibras" (o, más comúnmente, "fibróptica") en 1956.

A partir de entonces, la fibróptica se ha desarrollado como una rama independiente de las comunicaciones. El potencial de utilizar ondas ópticas guiadas fue evidente para quienes trabajaban en telecomunicaciones. Este hecho, junto con el advenimiento de los dispositivos emisores de luz de heterounión a mediados de la década de 1960 (en la obra de Panish⁴ se presenta una revisión de los primeros avances), alentó la investigación en laboratorio de las aplicaciones a la telefonía. Sin embargo, debido a la atenuación en la fibra (por lo general en el intervalo de los 1 000 dB/km), la

aplicación práctica de las fibras ópticas para las comunicaciones a larga distancia era más bien dudosa. Durante el período de 1966 a 1969, Kao y colaboradores⁵⁻⁸ de los Standard Telephone Laboratories señalaron que si se utilizaban materiales más puros sería posible reducir sustancialmente las pérdidas ópticas. Además, se dedicaron tenazmente al desarrollo de técnicas de medición para vidrios de ultrabaja atenuación. Esto estimuló la realización de mayor trabajo en diferentes laboratorios de todo el mundo, incluyendo British Post Office, Telefunken, Siemens Halske, Nippon Electric, Nippon Sheet Glass, Bell Telephone Laboratories y Corning Glass Works⁹⁻¹². El crédito de haber hecho posible la utilización práctica de las fibras ópticas en las comunicaciones pertenece a Kapron y colaboradores¹² de Corning: en 1970 anunciaron haber reducido la atenuación en la fibra a menos de 20 dB/km. Esto propició que la industria redoblara sus esfuerzos competitivos.

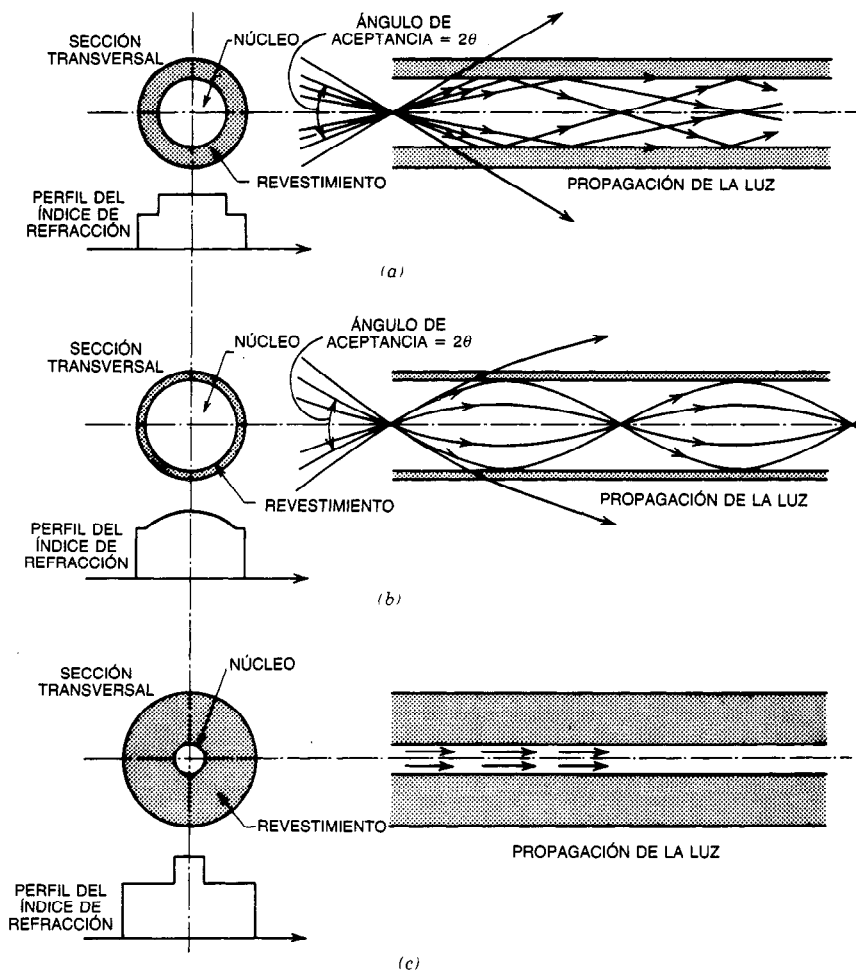


Fig. 48-90. Representación esquemática de fibras de uso común: a) fibra de índice escalonado, de 20 a 100 MHz-km; b) fibra de índice graduado, de 400 a 1 000 MHz-km; c) fibra monomodal, hasta 40 GHz-km.

48.4.2 Tipos de fibras y características de transmisión

Tipos de fibras

Desde el punto de vista histórico y de la propagación de la luz, las fibras ópticas pueden clasificarse en tres categorías:

1. Haces de fibras.
2. Fibras multimodales (de índice escalonado o índice graduado).
3. Fibras monomodales.

El haz de fibras, como el nombre implica, es un atado de fibras individuales ligeramente adheridas entre sí. Se utilizó prácticamente en todos los primeros experimentos. Sin embargo, cada fibra constituyente representa un núcleo con índice de refracción (n_1) relativamente alto rodeado de un revestimiento de menor índice de refracción (n_2). Cuando aparecieron mejores componentes optoelectrónicos y dispositivos de acoplamiento, se volvió preferible una sola fibra para las telecomunicaciones, debido a que con ella se eliminaban la atenuación intersticial y los problemas de dispersión relativa. Actualmente estas fibras simples se denominan también fibras multimodales, ya que con todas ellas son posibles una multitud de modos ópticos de propagación. Como se muestra en la figura 48-90, una fibra multimodal puede ser de índice escalonado o de índice graduado. La fibra de índice escalonado tiene un núcleo con índice de refracción constante⁸; por otra parte, el núcleo de una fibra con índice graduado varía de manera aproximadamente parabólica a través de su diámetro^{10,11}.

Según la ley de Snell, los rayos de luz que inciden en la frontera entre el núcleo y el revestimiento a un ángulo menor que el ángulo crítico θ se reflejan internamente por completo, como se muestra en la figura 48-90. Por tanto, existe un ángulo cónico de aceptación 2θ para cualquier fibra dada. Este ángulo cónico también está relacionado con la abertura numérica (que se analizará a continuación). Como se muestra en la figura 48-90, en las condiciones más perfectas en la fibra con índice graduado los rayos de luz aceptados se vuelven a enfocar de manera periódica a fin de preservar la relación de fase relativa entre los rayos, mientras que en la fibra con índice escalonado los rayos tienden a defasarse al cabo de una distancia. Por consiguiente, la fibra con índice graduado presenta menor dispersión modal y es preferida para las telecomunicaciones. La dispersión modal suele especificarse en productos de ancho de banda por distancia (MHz-km) o de tiempo por distancia unitaria (ns/km); en la figura 48-90 se presentan valores comunes. Una fibra que actualmente ha adquirido importancia es la monomodal (fig. 48-90c), en la que el núcleo está tan restringido (de 3 a 5 μm de diámetro) que esencialmente sólo los rayos axiales pueden propagarse. Esta fibra, que presenta la menor dispersión modal, es la preferida para las comunicaciones a larga distancia¹³. Desafortunadamente, debido a su pequeño diámetro, el alineamiento de las fibras monomodales es difícil, de forma que los co-

nectores desmontables de baja atenuación y los empalmes de fibra no son tan fáciles de obtener como los correspondientes a las fibras multimodales.

Abertura numérica (NA)

Un parámetro importante para la caracterización de las fibras es la abertura numérica (NA, de *numerical aperture*), una medida de la aceptación de luz en la fibra. Para las fibras de índice escalonado y monomodales (que básicamente son de índice escalonado), la NA se define como¹⁴:

$$NA (\text{índice escalonado}) = n_1 \sin \theta = \sqrt{n_1^2 - n_2^2} \quad (48-126)$$

donde θ = ángulo semicónico interno de aceptación

n_1 = índice del núcleo

n_2 = índice del revestimiento

Para una fibra de índice graduado, la NA es función del radio r al cual se inyecta la luz dentro de la fibra¹⁵:

$$NA(r) = n(r) \sin \theta = [n^2(r) - n_0^2]^{1/2} = \{2n_1^2[1 - (r/a)^2]\Delta\}^{1/2} \quad (48-127)$$

donde $n(r)$ es el perfil del índice de refracción como función de r , dado por $n_1[1 - \Delta(r/a)^2]$, y a es el radio del núcleo de índice graduado.

Para la inyección axial, es decir, para $r = 0$, la $NA(0)$ de una fibra de índice graduado se reduce a NA (índice escalonado).

Atenuación en la fibra y dispersión material

La atenuación de una fibra depende de la longitud de onda, y el valor límite es la atenuación por dispersión de Raleigh, cuya dependencia es de λ^{-4} . Sin embargo, las impurezas contenidas en una fibra (sobre todo radicales OH^- y diversos iones metálicos) contribuyen de forma significativa a los picos resonantes en la curva de atenuación en relación con la (frente a) longitud de onda, de lo cual se muestra un ejemplo en la figura 48-91. Las curvas de la figura son curvas compuestas que muestran los logros en cuanto a atenuación en las fibras. Al principio, la ventana para transmisión más importante estaba en el intervalo de 800 a 900 nm. Más tarde, sin embargo, a través del trabajo continuo en el proceso de control y otros aspectos, pudo disponerse comercialmente de fibras con "doble ventana" con baja atenuación en los intervalos de 800 a 900 nm y de 1.1 a 1.6 μm . Se ha informado que en laboratorio se han obtenido fibras con atenuaciones de 0.2 dB/km a 1.55 μm ¹⁶. Por lo general se ha afirmado que las fibras con esta baja atenuación deben elaborarse a partir de materiales iniciales ultrapuros (99.999%); sin embargo, experimentos posteriores indican que con materiales iniciales de menor pureza (99.17%) sería posible obtener una atenuación de 0.6 dB/km a 1.55 μm . Esto podría en el futuro reducir en gran medida el costo de manufactura de fibras de baja atenuación.

Las fibras también presentan propiedades de dispersión material que dependen de la longitud de onda. Para las fibras multimodales la dispersión material

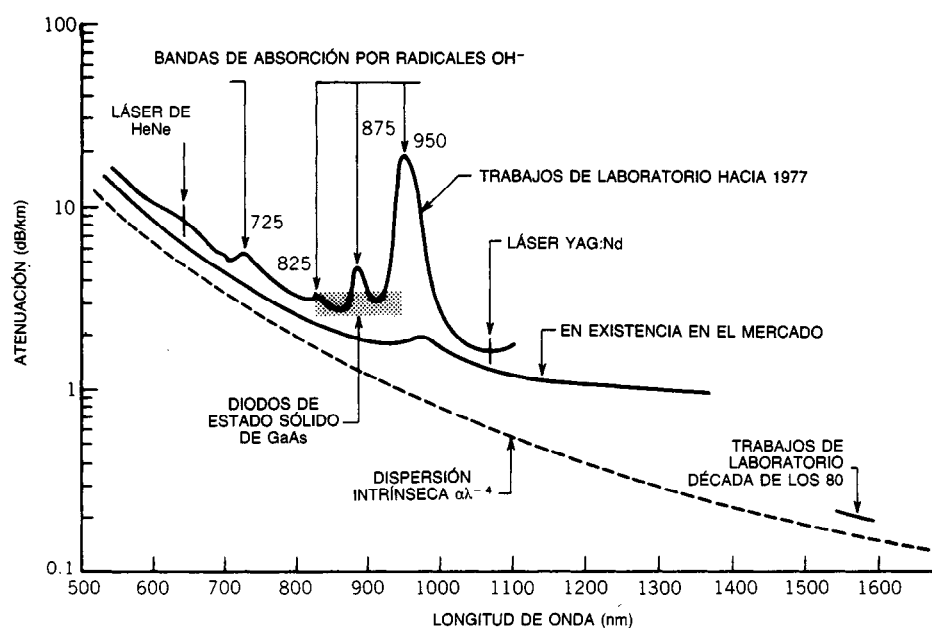


Fig. 48-91. Curvas de atenuación espectral compuesta de fibras ópticas de baja atenuación a la alta sílice.

suele ser pequeña en comparación con la dispersión modal. Por otra parte, para las fibras monomodales la dispersión material es el factor dominante¹⁷. Dado que las dispersiones material y modal varían en direcciones opuestas en un amplio intervalo de longitud de onda, es posible elegir material, razón de núcleo sobre revestimiento, etcétera, para una sola fibra, de forma que las dos dispersiones se cancelen esencialmente entre sí en un intervalo espectral de 1.35 a 1.67 μm ^{18,19}. Para tales fibras, el producto de ancho de banda para distancia es bastante mayor que 200 GHz-km para un ancho espectral de la fuente de menos de 5 nm²⁰.

Con base en este análisis de la atenuación y dispersión de la fibra, es claro que la tendencia futura será el empleo de fibras monomodales en la región de 1.3 a 1.6 μm para las telecomunicaciones a larga distancia. Sin embargo, por el momento y debido a la relativa inmadurez de fuentes, detectores, conectores, etc., las fibras monomodales en la ventana de 1.3 a 1.6 μm , muchos sistemas operacionales tienden todavía a permanecer en la región de 0.8 a 0.9 μm .

48.4.3 Técnicas de fabricación de fibras

A fin de minimizar la atenuación, toda la manufactura de la fibra debe iniciarse con materia prima muy pura. Naturalmente, para procesos diferentes se requieren diferentes materias primas. Estos procesos pueden agruparse en cinco categorías:

1. Método del crisol doble.
2. Oxidación externa en fase de vapor.
3. Deposición de vapor química modificada.
4. Deposición de vapor química por plasma.
5. Deposición axial en fase de vapor.

El método del crisol doble

Los materiales iniciales para este método son los vidrios altamente purificados, con impurezas de metal de transición totales del orden de 10 partes por mil millones. Los vidrios del núcleo y del revestimiento se funden en crisoles concéntricos; las pastas fundidas se extraen (al mismo tiempo que se las estira) por las boquillas situadas en la parte inferior de los crisoles, de modo que se forma una fibra revestida. En la figura 48-92a se muestra esquemáticamente este método. Durante los procesos de fusión y estirado deben tomarse precauciones a fin de evitar la contaminación por el ambiente o los crisoles. Para fibras de índice graduado, el perfil del índice se obtiene permitiendo la difusión de iones móviles a través de la interficie núcleo-revestimiento. En general, este método es idóneo para el estirado rápido de fibras con elevada abertura numérica²¹.

Oxidación externa en fase de vapor (OVPO)

En el método OVPO (de *outside vapor-phase oxidation*), reaccionantes en forma de vapor de alta pureza (como SiCl_4 , GeCl_4 , BCl_3O_2 , etc.) se queman a fin de obtener pequeñas partículas de vidrio de composición adecuada. Estas partículas, denominadas hollín, se depositan sublimadas sobre un mandril giratorio que también se desplaza axialmente adelante y atrás. De esta forma el hollín se deposita sobre el mandril, capa sobre capa, hasta constituir una preforma cilíndrica porosa que luego se retira del mandril. Según el material inicial y el control empleados, es posible elaborar preformas para fibras de índice escalonado o de índice graduado. Primero se sinterizan las preformas con objeto de formar "lingotes" de vidrio sólido, que luego se tratan con cloro para eliminar los iones hidroxilo

(OH⁻), que dan a las fibras picos de elevada absorción espectral. Luego, el lingote de vidrio se calienta en un horno hasta que adquiere la viscosidad adecuada y se estira en una fibra continua. Durante el proceso de estirado, el hueco central del lingote (que queda al retirarlo del mandril) se hace desaparecer a fin de obtener una fibra sólida. En la figura 48-92b se muestra este método en forma esquemática.

Con este proceso se producen comercialmente fibras continuas altamente concéntricas que presentan atenuaciones menores de 3 dB/km y productos de ancho de banda por longitud de 1 GHz-km. De uno de los lingotes es posible obtener fácilmente hasta 10 km de fibra con 125 μ m de diámetro.

Deposición química de vapor modificado (MCVD)

La MCVD (de *modified chemical vapor deposition*) es casi la operación inversa de la OVPO, en el sentido de

que se provoca que los vapores iniciales reaccionen dentro de un tubo de cuarzo de alta pureza. El hollín así obtenido se deposita dentro del tubo de cuarzo, que se hace girar a fin de asegurar una deposición uniforme. Al mismo tiempo, un soplete de varias llamas se desplaza a lo largo del tubo, de modo que lo caliente a una temperatura elevada, a fin de sinterizar el hollín en una capa vítrea (fig. 48-92c). De nuevo, es posible obtener diferentes perfiles del núcleo y revestimiento mediante la elección apropiada de los vapores y el control de la deposición del hollín. Una vez realizadas la sinterización y la deposición, tanto el tubo de cuarzo como su contenido se colapsan formando un lingote macizo que después se estira para obtener la fibra.

Con el empleo de este método se han obtenido fibras que varían desde multimodales con alta abertura numérica hasta monomodales con ultrabaja atenuación y dispersión cero²³.

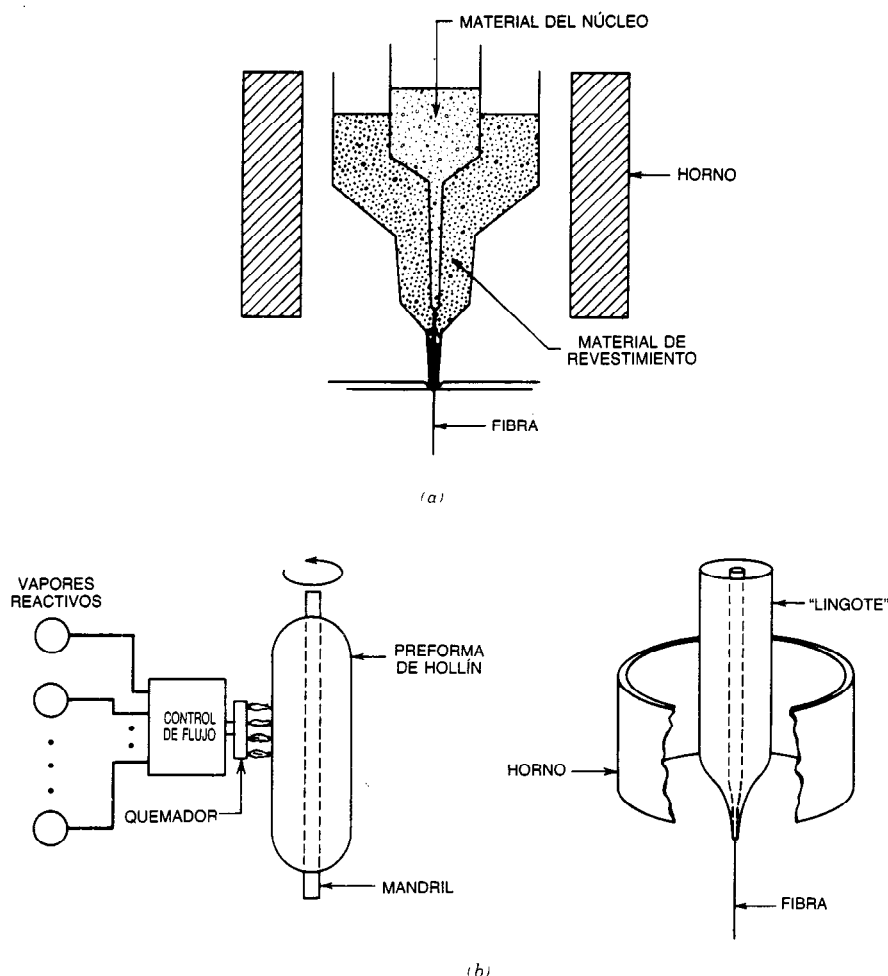
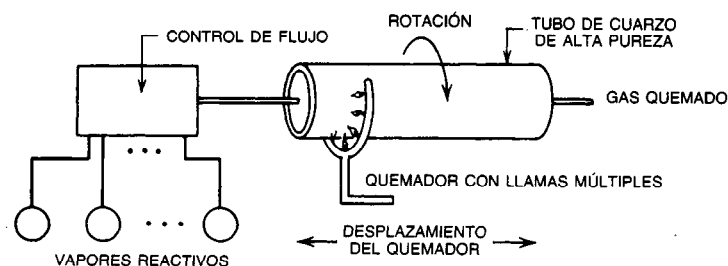
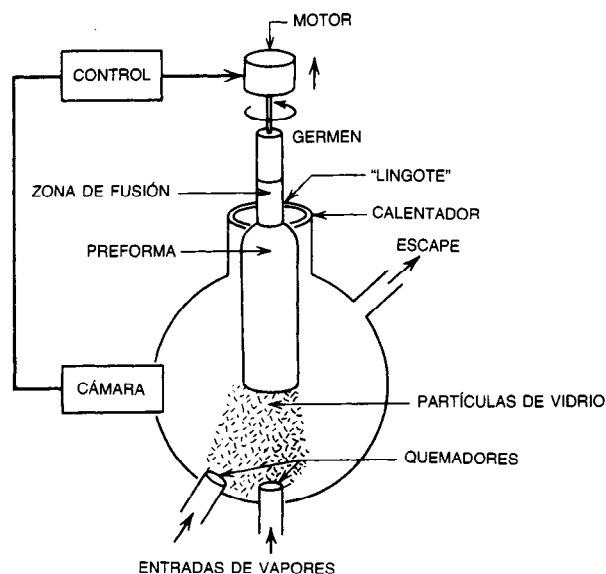


Fig. 48-92. Representación esquemática de algunos procesos de estirado de fibra: a) proceso del crisol doble (según N. S. Kapany. Cortesía de Academic Press); b) proceso de oxidación externa en fase de vapor: elaboración de la preforma (izq.) y estirado de la fibra (der.) (según P. C. Schultz, IEEE).



(c)



(d)

Fig. 48-92 (cont.). Representación esquemática de algunos procesos de estirado de fibra: c) deposición química de vapor modificada (según J. B. MacChesney, IEEE); d) deposición axial en fase de vapor (según T. Izawa y colaboradores, IEEE).

Deposición química de vapor por plasma (PCVD)

En principio, la PCVD (de *plasma chemical vapor deposition*) es bastante semejante a la MCVD, con la única diferencia de que la reacción de los vapores y la deposición del hollín se llevan a cabo por efecto de un plasma de RF o microondas y no de una llama. Aunque originalmente este proceso se considera "lento", trabajos posteriores efectuados en Philips han demostrado que es factible la producción comercial²⁴. Las fibras así producidas tienen típicamente una atenuación menor que 4 dB/km y un producto de ancho de banda por longitud de 0.5 a 1 GHz-km.

Deposición axial en fase de vapor (VAD)

El proceso VAD (de *vapor-phase axial deposition*) para producir preformas se parece más bien al desarrollo de cristales: se utiliza una varilla inicial de sílice

como germen, sobre el cual se deposita el hollín para la obtención de la preforma porosa. El hollín se genera al quemar los vapores crudos en un recipiente cerrado, mientras la varilla de sílice se hace girar y se desplaza hacia arriba. El calentamiento de la región cercana a la parte superior de la cámara de reacción cambia la preforma porosa en un lingote vítreo (fig. 48-92d). A fin de preservar la forma geométrica cilíndrica y un diámetro constante, la interficie entre la preforma y el vapor debe fijarse en cuanto a posición y mantenerse en exactitud. Esto suele realizarse mediante la observación a través de una cámara de TV y complejos circuitos de control electrónico²⁵. Con este proceso se evita el hueco central requerido en los procesos OVPO, MCVD y PCVD, y con él se han obtenido fibras de alta abertura numérica y monomodales de ultrabaja atenuación.

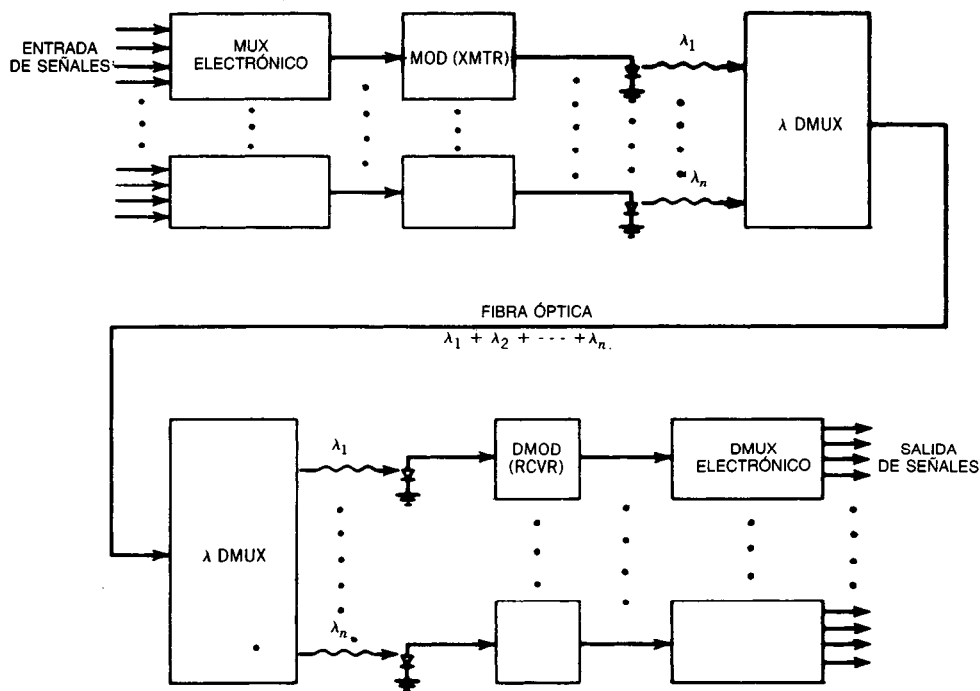


Fig. 48-93. Configuración general de un sistema de comunicaciones. MUX = multicanalizador; MOD (XMTR) = transmisor-modulador electrónico; λ MUX = multicanalizadores de longitud de onda; λ DMUX = desmulticanalizadores de longitud de onda; DMOD (RCVR) = receptor-desmodulador electrónico; DMUX = desmulticanalizador.

48.4.4 Componentes y sistemas de transmisión por fibras ópticas

Consideraciones del sistema

Debido a sus propiedades de baja atenuación, baja dispersión y baja detección, las fibras ópticas son ideales para transmisiones a larga distancia por banda amplia. Además, debido a que la propagación con bajas atenuación y dispersión es posible en un amplio intervalo de longitud de onda (de hecho, varios intervalos de longitudes de onda para las nuevas fibras), y a que los espectros de emisión de las fuentes luminosas pueden hacerse muy estrechos (p. ej., diodos láser con ancho de línea del orden de los 2 nm), es posible multicanalizar varias longitudes de onda en una sola fibra, de modo que cada longitud de onda pueda utilizar toda su propia capacidad de canal de comunicación. Diversas técnicas de multicanalización por división de longitud de onda (WDM, de *wavelength division multiplexing*) están recibiendo cada vez mayor atención en el laboratorio, y se están utilizando en forma creciente en sistemas prototipo. Por tanto, el sistema de fibras ópticas más general es como el que se muestra en la figura 48-93, en el que se han multicanalizado electrónicamente, de manera conjunta, varias señales de ancho de banda reducido (p. ej., canales de televisión o enlaces de datos de vídeo) a fin de formar un canal de comunicación óptica con un solo constituyente, de los cuales algunos se multicanalizan ópticamente a su vez para formar el canal óptico final de varias longitudes

de onda, que se transmite por una sola fibra. En el extremo receptor ocurre lo contrario y, por último, se recupera cada señal de ancho de banda reducido. Estas señales individuales pueden ser digitales o analógicas, y sus anchos de banda pueden variar de unos cuantos kilohertz a decenas de megahertz.

Con base en la figura 48-93, es evidente que para el diseño de un sistema de transmisión por fibras ópticas eficiente primero es necesario agrupar de manera racional las señales individuales a fin de minimizar los problemas en la multicanalización electrónica. A continuación, la señal electrónica compleja así obtenida se utiliza para modular la fuente luminosa (p. ej., la longitud de onda λ_1) en el transmisor. Aun cuando la multicanalización electrónica, la modulación, la desmodulación, etc., forman parte de un sistema de fibras ópticas, son por naturaleza electrónicas y se analizan en la sección 48.1. En esta sección el interés recae en los componentes optoelectrónicos y ópticos.

Fuentes luminosas: diodos emisores de luz y láseres

Aunque en teoría es posible utilizar muchos tipos de fuentes luminosas en los sistemas de comunicación por fibras ópticas, solamente los dispositivos semiconductores que radian en las regiones de 0.8 a 0.9 y de 1.06 a 1.55 μm revisten importancia práctica, debido a su compactidad y sencillez, así como facilidad de modulación y acoplamiento a las ventanas de transmisión de la fibra. Por consiguiente, en esta sección se presentan tales dispositivos con algún detalle.

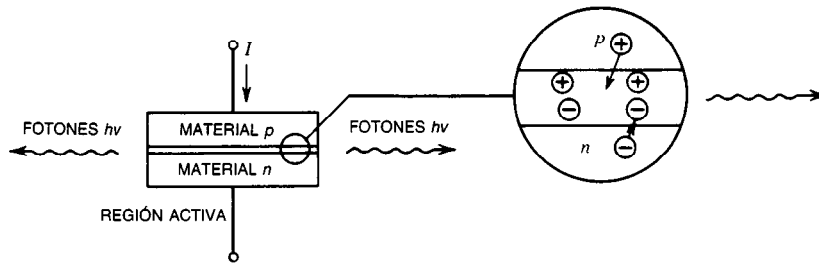


Fig. 48-94. Diagrama simplificado de una fuente semiconductor emisora de luz.

Básicamente, todos los emisores considerados aquí están contruidos de materiales impurificados, controlados diferencialmente durante su deposición, a fin de formar una unión entre materiales de tipo *p* y de tipo *n* (unión *pn*). Por lo general, el grosor de la unión es del orden de $1\ \mu\text{m}$ y su ancho es de 10 a $15\ \mu\text{m}$. En el funcionamiento, se hace pasar una corriente eléctrica por el dispositivo, de modo que se inyectan electrones y huecos a la unión. Estos portadores se combinan radiativamente en la unión para emitir fotones. En la figura 48-94 se muestra un diagrama de este proceso. Cuando la emisión de los fotones es incoherente, el dispositivo funciona como un diodo emisor de luz (LED, de *light-emitting diode*). Cuando las caras anterior y posterior del dispositivo están bien cortadas, de modo que el dispositivo forma su propia cavidad de Fabry-Perot, la acción láser se produce si la corriente de activación es suficientemente alta. Mediante el empleo de diferentes materiales anfitriones, impurificadores controlados y concentraciones de estos últimos es posible ajustar de manera precisa el emisor a su longitud de onda de radiación.

La figura 48-94 sugiere con claridad que al variar la magnitud de la corriente de activación I cambia en forma correspondiente la potencia de salida óptica. De hecho esto es cierto, como se muestra en la figura 48-95 para un diodo láser. A bajas corrientes de activa-

ción no existe suficiente retroalimentación positiva para producir la acción láser, y el dispositivo se comporta como un LED. A medida que se incrementa la corriente y se rebasa el umbral, ocurre efecto láser y la curva de salida de potencia óptica en función de la corriente de activación es esencialmente lineal. A una mayor corriente de activación se alcanza la saturación. Para un LED se obtiene una curva parecida, excepto que no se observa una inflexión pronunciada correspondiente a la acción láser. Es obvio que para aplicaciones en sistemas analógicos, el emisor debe estar polarizado en el centro de la región lineal y el giro en la señal de modulación debe estar bajo el tramo lineal. Para aplicaciones digitales, la polarización puede efectuarse en el corte, y se permiten corrientes de activación más elevadas.

Dado que los LED son dispositivos no coherentes, tienen un gran ancho espectral de salida (p. ej., de 50 a 60 nm en los 820 nm), amplios ángulos de emisión y menores anchos de banda de modulación (~ 150 MHz). Por tanto, suelen utilizarse en sistemas de corta distancia y baja velocidad de transmisión de datos con fibras de gran abertura numérica. Los láseres, que son coherentes, presentan reducidos anchos de banda espectrales (p. ej., de 2 a 4 nm en los 820 nm), ángulos de emisión estrechos y considerables anchos de banda de modulación (de varios GHz), y son más adecuados

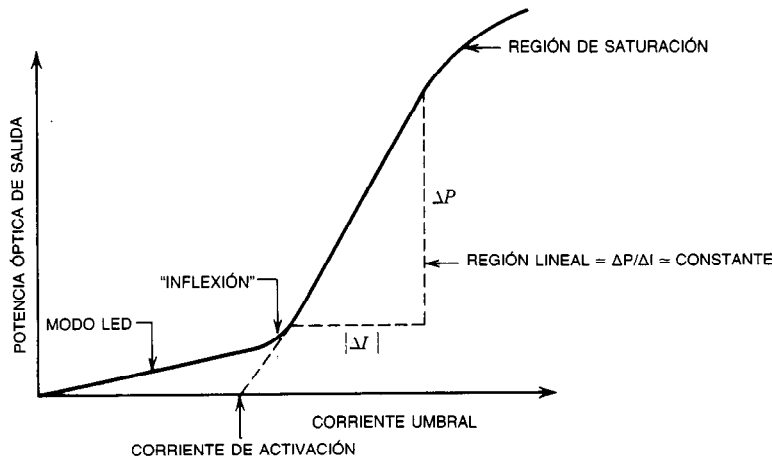


Fig. 48-95. Curva típica de la potencia en función de la corriente de activación para un láser de semiconductor.

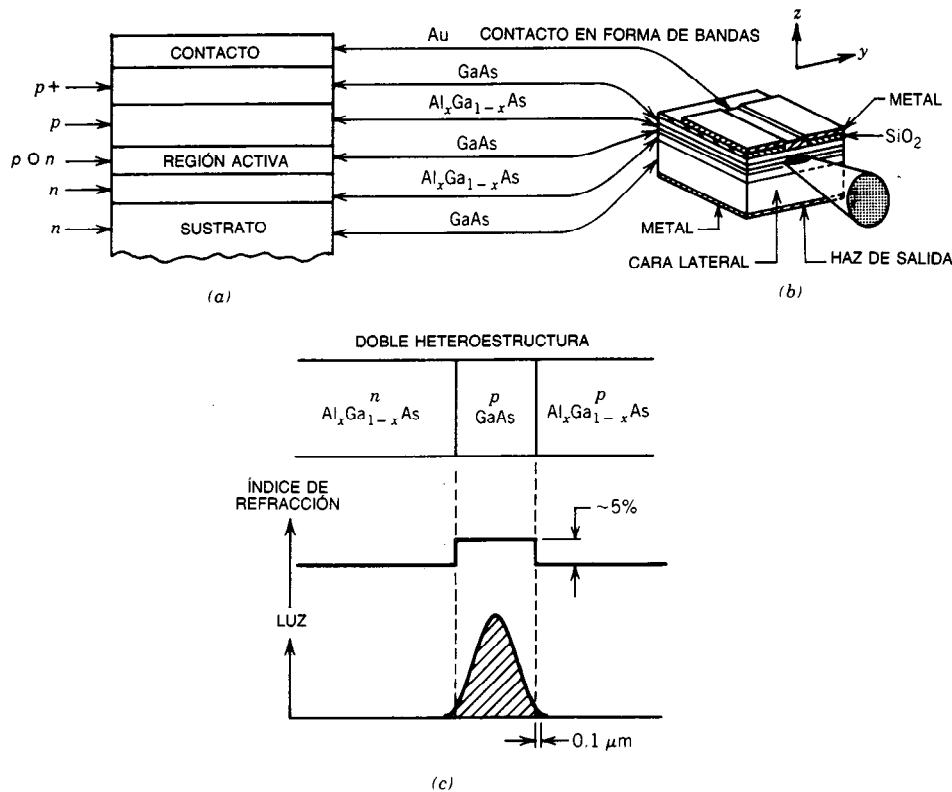


Fig. 48-96. Representación esquemática de un dispositivo de doble heteroestructura (DH) configurado en franjas: a) composición del material (cortesía de M. B. Panish); b) configuración del dispositivo (según T. P. Lee y colaboradores, IEEE); c) confinamiento óptico (cortesía de M. B. Panish).

para los sistemas de larga distancia y elevada velocidad de transmisión de datos en los que se emplean fibras de núcleo pequeño.

Desde la aparición de los semiconductores emisores a principios de la década de 1960, varios investigadores como Burrus, Kressel, Lee, Paoli y Ripper, por mencionar solamente unos cuantos, han realizado importantes trabajos. De los sistemas de materiales estudiados, tal vez el mejor estudiado es el de $\text{Ga}_x\text{Al}_{1-x}\text{As}$ sobre un sustrato de GaAs. Mediante el ajuste de los agentes impurificadores, este sistema emite en la región de 750 a 920 nm²⁶. Para la región de 1.06 a 1.55 μm se requieren sistemas con menor energía en la banda de transición; actualmente se está estudiando intensamente el sistema cristalino de doble mezcla $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ ²⁷.

El desarrollo de mayor importancia en toda el área de los semiconductores emisores de luz es tal vez la invención del dispositivo de doble heteroestructura (DH)²⁸⁻³⁰, tipificado por el sistema de arseniuro de galio-aluminio ($\text{Al}_x\text{Ga}_{1-x}\text{As}$ -GaAs- $\text{Al}_x\text{Ga}_{1-x}\text{As}$) que se muestra esquemáticamente en la figura 48-96. La doble heteroestructura deriva su nombre del hecho de que existen dos perfiles diferentes de impurificación (*p-p* y *n-n*) en cada lado de la región de emisión de luz. La disposición adecuada de los grosores de la capa

y de los niveles de impurificación permite obtener energías de la banda de transición e índices de refracción que den por resultado conducción y confinamiento efectivos en la dirección *z* (en la región activa GaAs). Además, si el contacto con el emisor se efectúa en una configuración por franjas, de modo que se restrinja aún más la corriente inyectada en la dimensión *y*, el resultado es que se requiere una corriente umbral muy baja para el efecto láser. La invención y mejora de tales dispositivos DH con configuración en franjas hizo posible la acción láser de ondas continuas (CW, de *continuous wave*) a temperatura ambiente, y estableció el empleo de dispositivos semiconductores como fuentes luminosas para los sistemas de comunicación por fibras ópticas. Hasta la fecha, la configuración DH en bandas sigue siendo la base de los láseres comerciales.

Los LED modernos se construyen siguiendo una línea semejante, utilizando estructuras DH. Sin embargo, dado que para su funcionamiento no se requiere el resonador de Fabry-Perot, la cara emisora puede configurarse a voluntad para lograr un acoplamiento óptico mejor. A este fin, el diodo de Burrus³¹, que se muestra en la figura 48-97, sigue constituyendo la mejor base para los LED comerciales. En tal diodo, un hemisferio se graba bastante adentro del sustrato a fin de reducir la absorción; se coloca una sola fibra cerca

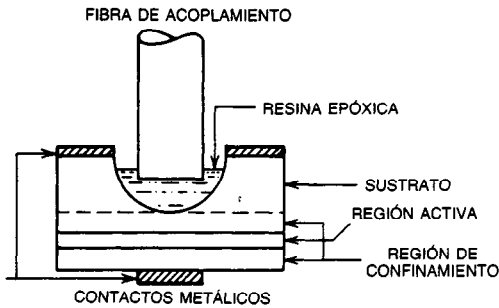


Fig. 48-97. Representación esquemática de un diodo Burrus (según S. E. Miller y colaboradores, IEEE).

del área emisora y se adhiere de forma permanente cementada con una resina epóxica para acoplamiento de índices. En la figura también se muestran con claridad la estructura DH y la configuración en franjas. Se han propuesto e intentado muchas otras variantes, entre ellas las formas planar y en domo³²⁻³⁴.

Fotodetectores

Como en el caso de los emisores, los fotodetectores más adecuados para los sistemas de comunicación por fibras ópticas son dispositivos semiconductores, como los fotodiodos PIN (de P-Intrínseco-N) y los fotodiodos de avalancha (APD, de *avalanche photodiodes*). Para el funcionamiento por debajo de la región de 1.06 μm , los dispositivos de silicio constituyen la mejor elección; para funcionamiento en la región de 1 a 1.6 μm suelen utilizarse otros materiales, tales como el germanio y arseniuro-fosforo de indio-galio. Por lo general, dado que la salida de una fibra es multimodal (o por lo menos tiende a cambiar los modos), la detección heterodina es impráctica. Por tanto, en los esquemas prácticos, todos los detectores se utilizan en el modo directo de detección (conteo de fotones)^{26,35,36}.

En la figura 48-98 se ilustra de manera esquemática el funcionamiento de un diodo PIN, en la que se muestra un fotón absorbido para generar un par de electrón-hueco, que después se separa en la región intrínseca agotada. La gran polarización inversa que suele utilizarse (p. ej., de 10 a 90 V) crea un gran campo eléctrico en la región de deriva, reduciendo con ello el tiempo de deriva y la capacitancia de la unión, que son dos factores esenciales para el funcionamiento a alta velocidad. La corriente externa i se induce en la carga R_L mientras los portadores están a la deriva en la región de agotamiento. Por tanto, i es linealmente proporcional a la intensidad incidente hasta que se alcanza la saturación. Para un diodo PIN de área pequeña a elevada polarización, es fácil obtener el funcionamiento sobre 1 GHz. La principal contribución al ruido en un diodo PIN de silicio es el ruido de perdigoneo debido a la corriente de oscuridad y a la corriente de fuga superficial. Sin embargo, en una situación normal domina el ruido térmico debido a R_L y a la primera etapa de amplificación.

En un fotodiodo de avalancha (APD), la unión pn^+ está bastante polarizada inversamente, como se mues-

tra en la figura 48-99, por lo que los portadores inducidos por fotones crean nuevos pares electrón-hueco mediante ionización por impacto. Los portadores secundarios continúan el proceso de ionización, dando por resultado multiplicación en avalancha. Por consiguiente, el APD combina de manera efectiva el conteo de fotones con la ganancia interna y es de utilidad en situaciones con poca energía recibida.

Sin embargo, dado que la multiplicación en avalancha responde esencialmente a un proceso estadístico, existe una fluctuación básica en la ganancia de corriente, de modo que el APD es fundamentalmente más ruidoso que un diodo PIN. Este ruido agregado suele especificarse mediante un factor de exceso de ruido $F(M)$ definido como:

$$\frac{d}{df} \langle i^2 \rangle = 2qI \langle M \rangle^2 F(M) \quad (48-128)$$

donde $d/df \langle i^2 \rangle$ = densidad espectral del ruido de la fotocorriente multiplicada

q = carga electrónica

I = corriente total del circuito

$\langle M \rangle^2$ = ganancia cuadrática media en avalancha

$F(M)$ también puede definirse en términos de $\langle M \rangle$ por medio de:

$$F(M) = \langle M \rangle_x \quad (48-129)$$

Para el silicio (Si), $x \approx 0.3$, mientras que para el germanio (Ge), $x \approx 1$.

En los primeros APD era difícil obtener uniformidad de la región de avalancha, de modo que con frecuencia ocurrían disrupciones prematuras, denominadas microplasmas. Esto podía reducir la ganancia e incrementar el ruido. Por medio de mejor procesamiento, empleo de anillos protectores, configuración del campo y otras técnicas, se ha obtenido ganancia uniforme en avalancha. Otras consideraciones que es necesario tomar en cuenta al utilizar el APD son mayor voltaje de polarización (~ 150 a 200 V) y sensibilidad a las variaciones de temperatura y voltaje. Actualmente muchos APD comerciales proporcionan medios

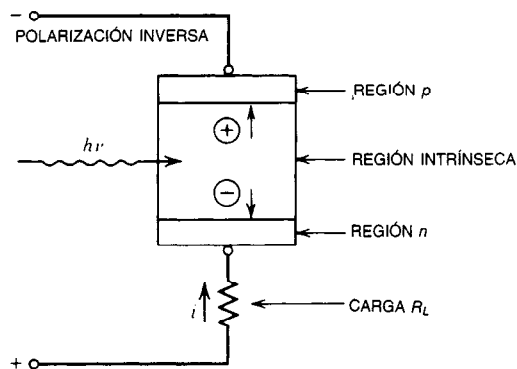


Fig. 48-98. Principio del funcionamiento de un diodo PIN (según R. G. Smith, IEEE).

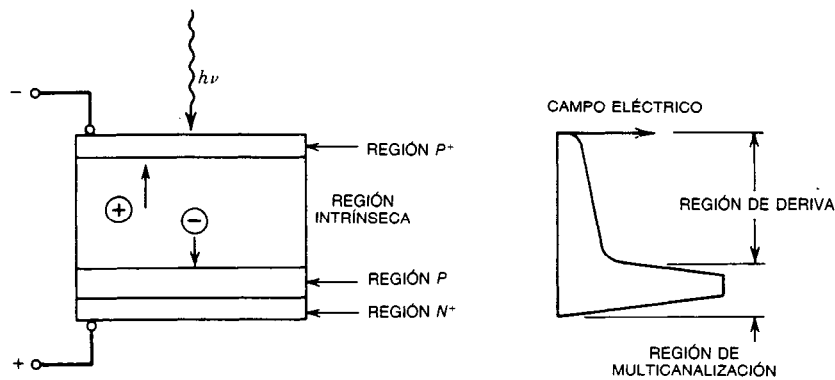


Fig. 48-99. Principio del funcionamiento de un fotodiodo de avalancha (según S. E. Miller y colaboradores, IEEE).

para el control de la retroalimentación de temperatura/voltaje, de forma que se minimizan tales efectos.

Conmutadores ópticos

La mayor parte de los conmutadores ópticos descritos en las publicaciones especializadas recientes son dispositivos ópticos integrados, en los cuales las dos guías de ondas ópticas se depositan como capas delgadas y se acoplan mediante un efecto electrónico durante la aplicación de un voltaje³⁷⁻⁴¹. Sin embargo, Soref y colaboradores⁴² describen un dispositivo de tres puertas elaborado a partir de un cristal de LiNbO_3 , que se muestra en la figura 48-100. La conmutación se efectúa induciendo las guías de ondas de los canales A y B (cada una de $220\text{ }\mu\text{m}$ de amplitud) en el cristal de LiNbO_3 por medio del efecto electrónico. En una posición de conmutación se aplica un voltaje V_a al electrodo superficial A, incrementando el índice de refracción bajo él y creando la guía de ondas de canal A. A fin de cambiar a la otra posición de conmutación se

aplica el voltaje V_b , creando el canal B y eliminando luz del canal A por medio de la acción de un acoplador direccional. Las atenuaciones totales inherentes son de unos 4 dB, y la longitud del conmutador es aproximadamente de 1.7 cm. Para utilizar este conmutador se requieren polarizadores y lentes (que no se muestran).

Los conmutadores ópticos que se encuentran con mayor facilidad son los de tipo acusticoóptico, en los que se emplea el principio de difracción de Bragg. Una onda acústica progresiva ("viajera"), excitada por una RF, se dirige hacia el cristal a fin de crear una variación periódica de su índice de refracción. Esto crea efectivamente una rejilla de difracción en el cristal. Cuando se hace incidir una luz láser sobre esta rejilla virtual en un ángulo adecuado, ocurre deflexión. La conmutación óptica de un haz de entrada entre dos ubicaciones de salida suele efectuarse mediante una conmutación electrónica entre dos osciladores cristalinos en el activador de RF. La atenuación por inserción suele ser de unos 2 dB.

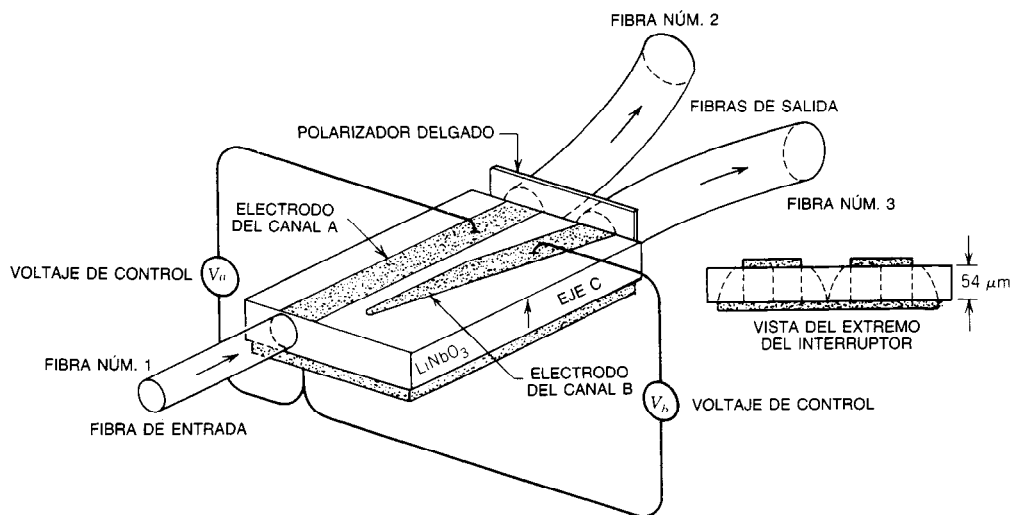


Fig. 48-100. Conmutador acoplador direccional multimodal de tres puertas (cortesía de R. A. Soref).

Aunque este tipo de conmutador tiende a ser voluminoso, es el más logrado en cuanto a tecnología. En la configuración real no se requiere polarizador y el empleo de lentes es opcional, dependiendo de los circuitos del sistema.

Multicanalizadores y desmulticanalizadores de longitud de onda

Los multicanalizadores de longitud de onda son por lo general de dos tipos: rejillas de difracción y filtros dieléctricos. En las publicaciones especializadas se ha informado sobre varios multicanalizadores de rejillas de difracción⁴³⁻⁴⁶; en ellos suele utilizarse una combinación lente-rejilla de difracción, como se muestra en la figura 48-101a. Los haces de luz de las fibras de entrada (λ_1 , λ_2 y λ_3) son colimados por las lentes sobre la rejilla de difracción soldada, y difractados por ésta conforme a sus longitudes de onda. Se vuelven a enfocar en la fibra de salida como la salida multicanalizada. En la práctica suele utilizarse una lente de varilla GRIN (de *graded refractive index*, índice de refracción graduado), debido a la ventaja de sus caras planas de entrada y salida, orientación coaxial de las fibras, fácil compensación de la abertura numérica, aberración mínima, facilidad para el montaje, etc. En la figura 48-101b se puede observar una configuración reciente en la que se utiliza una combinación lente de varilla GRIN/rejilla de difracción. La atenuación por inserción típica es de 2.5 dB aproximadamente, y el aislamiento entre canales separados 20 nm entre sí es de unos 24 dB.

También se ha investigado ampliamente el empleo de filtros dieléctricos de interferencia de capas múltiples como multicanalizadores y desmulticanalizadores⁴⁷⁻⁴⁹. Dado que los filtros dieléctricos suelen presentar una elevada sensibilidad a la polarización, por lo general se utilizan en configuraciones de incidencia casi normal junto con varillas GRIN, como se muestra en la figura 48-102. Aunque la opinión generalizada ha sido que los dispositivos de interferencia proporcionan baja transmisión y aislamiento deficiente, un dispositivo bien diseñado, tal como el que se muestra en la figura 48-102a, presenta una atenuación por inserción media de alrededor de 2 dB y aislamiento entre canales (separados 20 nm entre sí) de 34 dB aproximadamente⁴⁵.

En virtud de que estos dispositivos son recíprocos, un multicanalizador también puede utilizarse como desmulticanalizador. Sin embargo, es posible obtener mayor salida óptica en un desmulticanalizador si las fibras de salida tienen núcleos más grandes que el núcleo de la fibra de entrada.

Acopladores de fibras ópticas: direccionales, en T y en estrella

En un sistema con terminales múltiples (p. ej., un ducto de datos) se requieren varios acopladores ópticos. Tres de los más elementales se analizan aquí.

Acopladores direccionales. Un acoplador direccional es un dispositivo de tres puertas, que se muestra esquemáticamente en la figura 48-103a. Idealmente, la po-

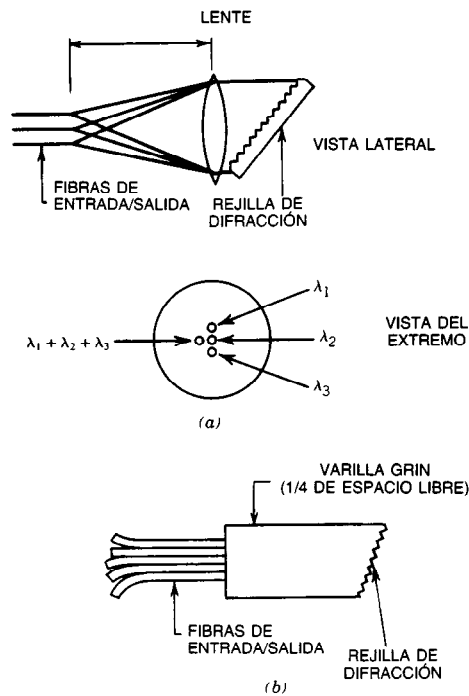


Fig. 48-101. Multicanalizador de longitud de onda en el que se utiliza una rejilla de difracción como elemento sensible a la longitud de onda: a) principio de funcionamiento (cortesía de W. J. Tomlison); b) reducción reciente a la práctica (cortesía de A. H. Fitch y colaboradores).

tencia de la puerta A se dirige sólo a la puerta C, la de ésta se dirige sólo a la puerta B, y existe aislamiento infinito entre A y B. Por tanto, tal acoplador se utiliza en un sistema dúplex. En la práctica, una construcción lograda es como la que se muestra en la figura 48-103b: dos fibras idénticas se retuercen juntas, se calientan y estiran a fin de obtener un "ahusamiento bicónico"⁵⁰. Los dispositivos comerciales construidos así presentan un aislamiento entre A y B mayor que 50 dB y una atenuación por inserción entre todas las puertas de 4 dB aproximadamente.

Acopladores en T. El acoplador en T (también denominado bifurcación) es asimismo un dispositivo de tres puertas, y se muestra en la figura 48-104. El principal flujo de potencia es de A a C, con una parte prescrita (la razón) de bifurcación proveniente de la puerta B. Aunque es posible obtener una construcción sencilla empleando una combinación fibra-filtro bastante semejante a la de un multicanalizador (fig. 48-102a), en dispositivos comerciales recientes vuelve a utilizarse la construcción ahusada bicónica. El exceso de atenuación $[(P_A - P_B - P_C)/P_A]$ de tal dispositivo suele ser de 1 dB aproximadamente.

Acopladores en estrella. Un acoplador en estrella es un dispositivo de n puertas en el que una señal de una fibra se distribuye uniformemente entre las demás fi-

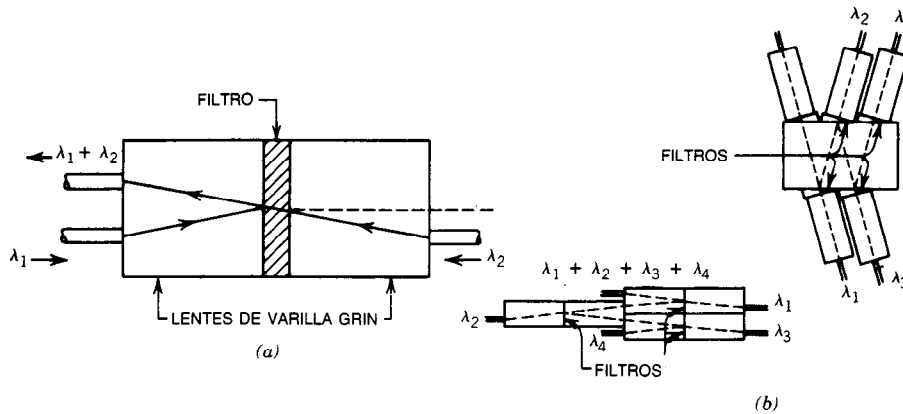


Fig. 48-102. Multicanalizador de longitud de onda en el que se utilizan filtros de interferencia dieléctrica de capas múltiples como elemento sensible a la longitud de onda: *a)* de una sola etapa (cortesía de A. H. Fitch y colaboradores); *b)* configuración en cascada (cortesía de W. J. Tomlison).

bras, como se muestra en la figura 48-105. Es decir, se permite que la potencia de entrada P_{ent} de cualquier fibra llene toda la varilla emisora, que por lo general es una gran fibra de índice escalonado, y después se refleja de regreso para iluminar de manera uniforme las fibras. Por tanto, idealmente cada fibra recibe P_{ent}/N de la señal. Los dispositivos prácticos tienen uniformidad entre puertos del orden de 1 a 2 dB, y un exceso de atenuación ($P_{ent} - \sum P_{sal}$) del orden de 4 dB.

Conectores y empalmes

Cuando dos fibras ópticas se conectan entre sí, pueden unirse en forma desmontable mediante el empleo de conectores o bien permanentemente en un empalme. Ya que las fibras suelen tener diámetros muy pequeños (angular y lineal), una consideración de importancia es el alineamiento. En la figura 48-106 se muestran las atenuaciones calculadas para fibras de índice escalonado con respecto a varias condiciones de desalineamiento y separación en los extremos, y se muestra que es posible obtener atenuaciones sustanciales si no se tiene cuidado al hacer las conexiones.

Los fabricantes actuales utilizan dos métodos básicos en sus diseños:

1. El empleo de dispositivos mecánicos (ranuras, casquillos de alineamiento, cojinetes de piedra dura [rubí], etc.) a fin de alinear el diámetro de las fibras.
2. El empleo de elementos ópticos (lentes, varillas Selfoc, lentes virtuales, etc.) a fin de obtener mayor diámetro, haz de salida colimado para ser recibido por el sistema óptico de acoplamiento de entrada.

En el primer método los factores de pérdida dominantes son el desplazamiento y la separación axiales, mientras que en el segundo básicamente se cambia el problema de desplazamiento axial por desalineamiento angular. Ya que el desalineamiento angular puede obtenerse con mayor facilidad si es posible mantener una

rigidez mecánica en una gran distancia, y ya que un haz con gran diámetro es menos sensible a inclusión de polvo, rayaduras, etc., el método óptico suele proporcionar resultados superiores, especialmente después de varios acoplamientos y desacoplamientos.

En la figura 48-107 se muestran los principios de dos diseños de uso actual. La figura 48-107*a* corresponde a un método de ranuras múltiples utilizado por los Bell Telephone Laboratories. Sobre una oblea de dióxido de silicio se graban con precisión 12 ranuras, entre dos obleas se embeben 12 fibras en una resina epóxica, y se apila un total de 12 capas (144 fibras). A continuación, el bloque así formado se pule para soldarlo a tope con un bloque de acoplamiento. Se han informado atenuaciones por inserción de 0.2 a 0.3 dB^{15, 51}. En la figura 48-107*b* se muestra el principio de un conector desmontable de tipo lente, disponible en el comercio. El corazón del conector es un elemento plástico moldeado, en el interior de cuyas cámaras cónicas sobresalen las fibras sin recubrir; se utiliza fluido óptico en las cámaras cónicas para la compensación y lentificación. A fin de minimizar la atenuación por inserción, los parámetros de la lente se eligen de modo que op-

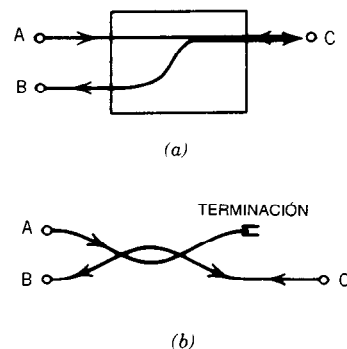


Fig. 48-103. Acoplador direccional: *a)* principio de funcionamiento; *b)* reducción a la práctica.

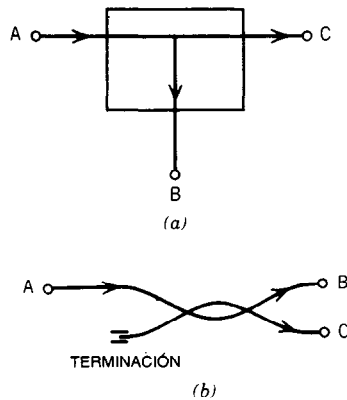


Fig. 48-104. Acoplador en T: a) principio de funcionamiento; b) reducción a la práctica.

timicen el acoplamiento de las fibras específicas, y la sección biconica se moldea con precisión para obtener alineamiento transversal. Por lo general es posible mantener la atenuación por inserción en aproximadamente 1 dB, incluso después de varios acoplamientos.

Actualmente es de uso corriente otra técnica, denominada de empalme por fusión. Las fibras por empalmar se desnudan, sus extremos se juntan a tope, y las fibras se alinean bajo un microscopio por medio de micromanipuladores. Después se aplica un arco eléctrico para fundir los extremos entre sí, formando un enlace permanente. Con la práctica se obtienen extremos cuadrados bien unidos y limpios; los empalmes por fusión suelen tener atenuación de 0.1 a 0.2 dB y son mecánicamente estables y resistentes. A fin de mantener la integridad mecánica después de la formación del empalme se requiere una estructura externa de apoyo.

48.4.5 Efectos de las radiaciones nucleares

Dado que las fibras ópticas se elaboran con sílice y silicatos, con varios impurificadores dosificados, no son intrínsecamente resistentes a las radiaciones ionizantes. Es decir, la radiación dañará o romperá enlaces entre átomos, y por tanto establecerá sitios para el atrapamiento de carga en la matriz vítrea. Estos sitios crean centros de color en la fibra, lo que incrementa bastante la atenuación. La cantidad total de daño aumenta con la dosis total recibida de forma altamente no lineal. Por fortuna tal daño no es permanente y por lo general se recupera con el tiempo, en especial a temperaturas elevadas. Los procesos de recocido son complicados y dependen no sólo de la composición de la fibra anfitriona y de la concentración de impurificadores, entre otros factores, sino también del proceso de estirado de la fibra y de los antecedentes previos de la irradiación. Aun cuando tal recocido es continuo en el tiempo, la razón de recuperación no es espectralmente uniforme debido a que la absorción en las distintas regiones espectrales puede ser provocada por diferentes defectos, de modo que los

mecanismos de recuperación pueden ser bastante diferentes. Se ha informado un efecto de recocido inducido ópticamente, lo que podría mejorar el endurecimiento de la fibra, en especial cuando no se dispone de temperatura de recocido.

A partir de este breve análisis, es bastante evidente que la respuesta de la fibra a un pulso de radiación difiere mucho de su respuesta a la radiación de estado estable. En el caso del pulso (transitorio), la atenuación de cresta transitoria depende de la velocidad de generación de daño (intensidad de dosis) y de la velocidad de recuperación continua, mientras que en el caso de estado estable el daño total es el resultado neto de dos procesos estables en desarrollo y, por tanto, sólo depende de la dosis total. Además, en la radiación transitoria, la emisión de pulsos de luz suele ocurrir con una amplitud de pulso generalmente igual a la propia de la radiación ionizante. Así, la respuesta transitoria de una fibra por lo general es como la que se muestra en la figura 48-108. El origen de esta emisión de luz no se ha dilucidado del todo: una escuela de pensamiento la atribuye a la emisión espontánea, mientras que las demás la atribuyen a la radiación de Cerenkov. En cualquier caso, la luz emitida tiende a estar en la región de la longitud de onda corta y es posible filtrarla eficazmente si el sistema transmite en la región del infrarrojo próximo (p. ej., a 820 nm).

Debido a los mecanismos de respuesta no lineal y a la intensa dependencia de la recuperación con respecto a parámetros medioambientales, la comparación de datos de prueba es difícil. Hablando estrictamente, la comparación puede efectuarse sólo si todos los parámetros son idénticos: esto incluye no solamente las condiciones de prueba y medición, sino también el almacenamiento y manipulación antes y después de la radiación. Sin embargo, con los diversos datos de prueba con que se cuenta en la actualidad es posible indicar las siguientes tendencias de las tres clases más conocidas de fibras [es decir, de sílice impurificada, de sílice revestida de polímeros (PCS, de *polymer clad silica*) y de vidrio de silicato impurificado] en el intervalo espectral de 0.8 a 0.9 μm .

Radiación de estado estable. Ésta suele realizarse utilizando una fuente de cobalto 60. A bajas dosis (menos de 1 000 rad si) las fibras de sílice impurificada tienden a presentar baja atenuación, y a altas dosis (1 000 a 10 000 rad si) son las fibras PCS las que suelen tener menor atenuación. Desafortunadamente, las fibras de silicato de vidrio suelen exhibir elevada atenuación a todos los niveles.

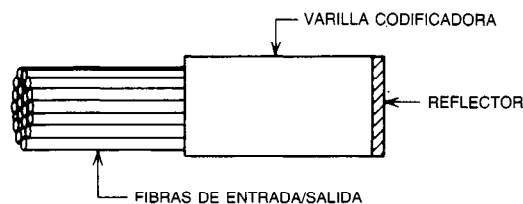


Fig. 48-105. Acoplador en estrella.

Respuesta a la radiación transitoria. Por lo general tales pruebas de radiación se realizan en un generador de pulsos de rayos X o en una fuente de electrones. De nuevo, las fibras PCS tienden a presentar baja atenuación transitoria de cresta, así como rápida recuperación a temperatura ambiente, seguidas casi siempre por las fibras de sílice impurificada.

A longitudes de onda mayores (p. ej., en el intervalo de 1.3 a 1.5 μm), las mediciones hasta la fecha indican que el endurecimiento contra la radiación es mejor sustancialmente para sílice impurificada, y más aún para el PCS. Por consiguiente, con la elección adecuada de longitud de onda, material para la fibra, impurificadores y concentración de impurificadores, procesamiento controlado de la fibra, temperatura de

operación y consideraciones de manejo y almacenamiento, las fibras ópticas pueden utilizarse para diversos ambientes transitorios y de estado estable, como es evidente a partir del gran número de sistemas prototipo para estaciones de energía nuclear e instalaciones militares.

Hasta el momento, el análisis se ha concentrado en la respuesta de las fibras ópticas, ya que son los elementos "nuevos" o "desconocidos" en un sistema de transmisión por fibras ópticas. El Department of Defense y el Department of Energy estadounidenses han publicado manuales para el diseño de circuitos electrónicos endurecidos contra las radiaciones nucleares y listas de piezas. De hecho, existe una gran cantidad de literatura especializada al respecto. Bá-

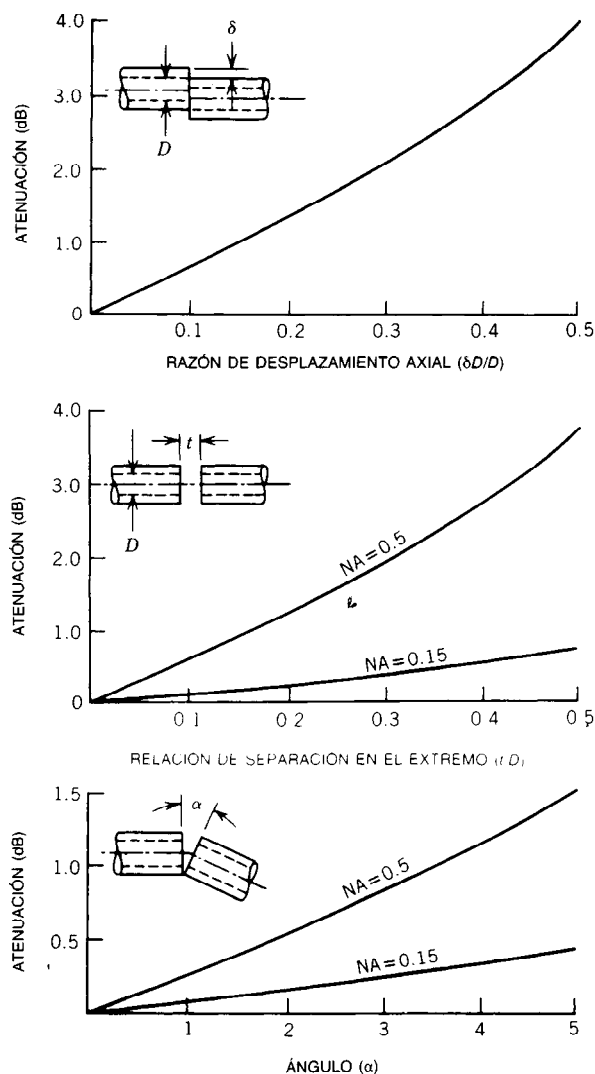


Fig. 48-106. Atenuaciones (pérdidas) calculadas de un empalme de fibra multimodal de índice escalonado, debidas a desalineamiento axial, separación entre los extremos y desnivel angular (cortesía de F. L. Thiel y colaboradores).

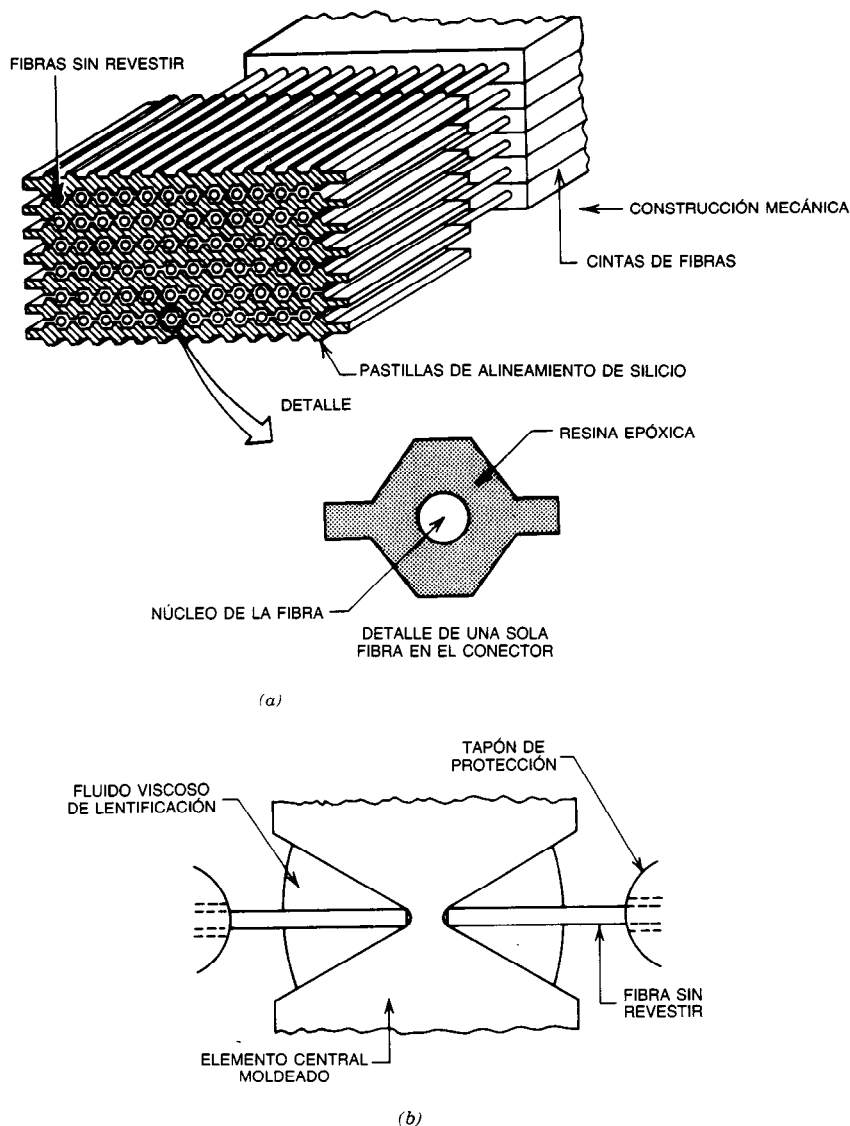


Fig. 48-107. Dos tipos de conectores de fibra óptica: a) bloque de ranuras (según J. F. Dalglish, IEEE); b) conector de una sola fibra (cortesía de The Deutsch Co.).

sicamente, los circuitos electrónicos se protegen siempre que es posible y se diseñan a fin de proporcionar la mínima respuesta del sistema a estímulos inducidos por radiación nuclear. Por consiguiente, suele sacrificarse sensibilidad por dureza nuclear. En el caso de una elevada radiación transitoria después de la cual la atenuación transitoria pico (cresta) de la fibra puede ser considerable, podría ser necesario especificar un tiempo de "paralización" aceptable para que el sistema se recupere de esta condición de elevada atenuación. Durante el tiempo de paralización los datos transmitidos pueden perderse totalmente o, por el contrario, almacenarse en una memoria endurecida en espera de la recuperación.

48.4.6 Ejemplos sencillos de diseño

Al diseñar un sistema de fibras ópticas, la primera consideración es el ancho de banda del sistema, debido a que determina la elección básica de la fuente óptica (LED frente a láser). La siguiente consideración es la longitud del enlace, ya que determina el tipo de fibra que debe utilizarse y qué atenuación global de la fibra debe esperarse. Esta última, junto con varias atenuaciones por acoplamiento, conexión, empalme y bifurcación, determina si debe utilizarse un diodo láser sin tomar en cuenta el ancho de banda, el tipo de detectores ópticos, y si son necesarios repetidores. Si existiera radiación nuclear, la elección de la fibra esta-

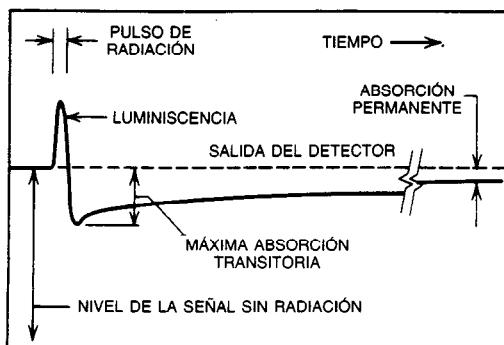


Fig. 48-108. Respuesta típica de una fibra óptica sometida a radiación por un pulso ionizante corto (cortesía de B. E. Kincaid y colaboradores).

rá determinada por los requisitos de endurecimiento a la radiación y por el tiempo de paralización permisible. En este caso, también se reducirá la sensibilidad del receptor.

Enlace de punto a punto, a larga distancia, con gran ancho de banda

En un enlace analógico de 200 MHz y 3.2 km, cuya fuente es el punto A, y con un sumidero conmutable en B o en C (fig. 48-109) el ancho de banda del sistema requiere el empleo de un buen diodo láser; la longitud y el ancho de banda requieren una buena fibra con índice graduado, cuyo producto de ancho de banda por longitud sea mayor de 600 MHz-km. Dado que las longitudes de las fibras comerciales suelen ser de 1 km, se indican tres empalmes. Para un buen diodo láser con cable flexible para conexión, podría esperarse una salida lineal modulada de 1 mW (0 dBm). El diodo también estará polarizado a una salida de CD de 1 mW. Por tanto, el presupuesto para el enlace será:

Potencia de la fuente	0 dBm
Atenuaciones del enlace	
3 conectores a 1 dB cada uno	3.0 dB
3 empalmes a 0.2 dB cada uno	0.6 dB
3.2 km de fibra a 4 dB/km	12.8 dB
1 interruptor óptico a 2 dB	2.0 dB
Atenuación por acoplamiento en el fotodiodo	1.0 dB
Envejecimiento, desalineamiento	3.0 dB
Atenuación total del enlace	22.4 dB
∴ Potencia incidente en el fotodiodo	- 22.4 dB

Utilizando un diodo PIN de sensibilidad igual a 0.5 mA/mW, la fotocorriente i del diodo es 2.9 μ A, lo que significa que el diodo PIN estará limitado por ruido de perdigoneo. Ya que la potencia del ruido de perdigoneo está dada por $2qi\Delta f$, mientras que la potencia de la señal está dada por i^2 , la razón de señal sobre ruido en el fotodiodo, $(S/N)_i$, es $i/2q\Delta f$. Ahora, $i = 2.9 \times 10^{-6}$ A, $q = 1.6 \times 10^{-19}$ C (carga del electrón), y $\Delta f = 200$ MHz = 2×10^8 Hz; por tanto:

$$(S/N)_i = \frac{2.9 \times 10^{-6}}{2 \times 1.6 \times 10^{-19} \times 2 \times 10^8} =$$

$$= \frac{2.9 \times 10^{-6}}{6.4 \times 10^{-11}} = 4.5 \times 10^4$$

o bien:

$$(S/N)_i = 46.5 \text{ dB}$$

Ésta no es la razón de señal sobre ruido del sistema, ya que el ruido del amplificador de la primera etapa (posdetector) también contribuye. La contribución exacta depende del esquema de detección, del diseño de los circuitos y de otros factores, y no puede asignarse arbitrariamente. Sin embargo, suponiendo un factor conservador de ruido igual a 5 dB para un amplificador de 23 dB, la razón de señal sobre ruido de salida $(S/N)_o$ se reduce a 41.5 dB, y ésta debe ser esencialmente la razón de señal sobre ruido del sistema.

Si el rendimiento requerido de señal sobre ruido es 35 dB, entonces el enlace actual tiene un margen de 6.5 dB, que es adecuado para la mayor parte de las aplicaciones. Para un requisito de mayor margen o razón de señal sobre ruido más alta debe utilizarse un APD, que proporcionará otra señal de 13 dB.

Enlace de terminales múltiples con ancho de banda intermedio

Considérese una situación de un ducto de datos con N terminales, en la cual una estación maestra transmite a las demás $(N - 1)$ estaciones esclavas. Por sencillez, se utiliza un acoplador en estrella; cada estación es equidistante con respecto a la estrella (fig. 48-110). Este ducto de datos transmitirá a unos 10 Mb/s, con codificación bifásica, a una separación de estrella a estación de 100 m o menos, y tendrá aproximadamente 32 terminales en el sistema. Estos parámetros indican que es posible utilizar un LED y una fibra con índice escalonado de gran abertura numérica, para obtener una entrada acoplada hacia el interior de la fibra de aproximadamente 100 μ W (-10 dBm) en promedio. El presupuesto de la potencia óptica será:

Fuente de potencia	-10 dBm
Atenuaciones del enlace entre la estación maestra y sólo una estación esclava	
4 conectores a 1 dB cada uno	4.0 dB
200 m de fibra a 4 dB/km	0.8 dB
Atenuación por división = $1/N = 1/32$	15.0 dB
Exceso de atenuación de la estrella	2.0 dB
Atenuación por acoplamiento en el fotodiodo	0.5 dB
Envejecimiento, desalineamiento	3.0 dB
Atenuación total de estación a estación	25.3 dB
∴ Potencia incidente en el fotodiodo	-35.3 dB/m

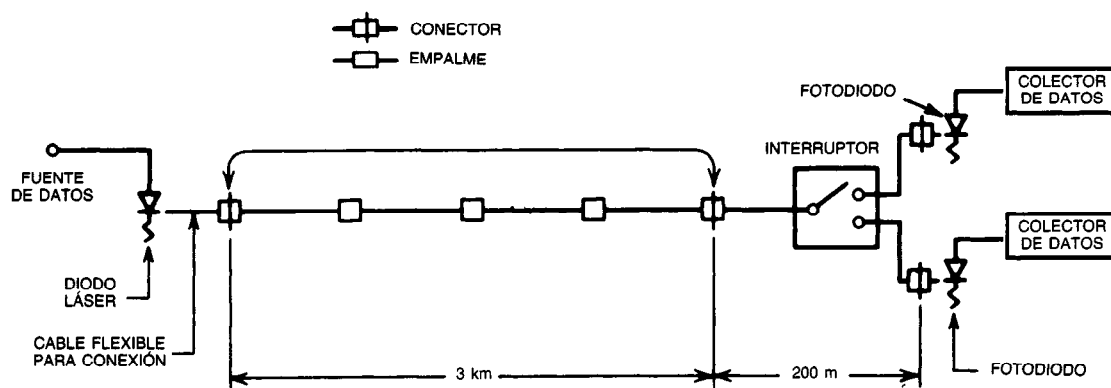


Fig. 48-109. Ejemplo de diseño: enlace de datos analógico.

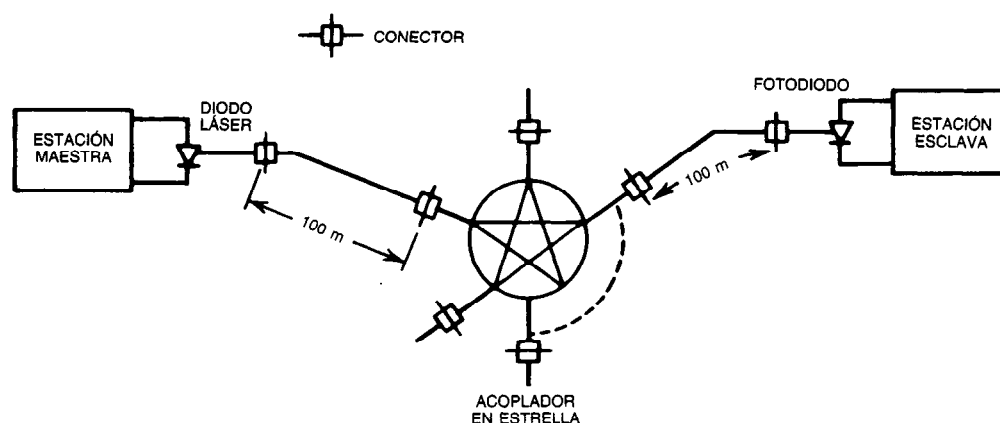


Fig. 48-110. Ejemplo de diseño: ducto de datos en el que se utiliza un acoplador en estrella.

Otra vez, utilizando el mismo diodo PIN con responsividad de 0.5 mA/mW, se encuentra que la fotocorriente es de alrededor de $.015 \mu\text{A}$, de modo que el detector sigue estando limitado por ruido de perdigoneo. El ancho de banda de una señal codificada bifásica de 10 Mb/s se estima conservadoramente en 40 MHz; es decir, $\Delta f = 4 \times 10^7 \text{ Hz}$:

$$\begin{aligned} \therefore (S/N)_i &= \frac{i}{2q\Delta f} = \frac{1.5 \times 10^{-7}}{2 \times 1.6 \times 10^{-19} \times 4 \times 10^7} = \\ &= \frac{1.5 \times 10^{-7}}{1.28 \times 10^{-11}} = 1.17 \times 10^4 \end{aligned}$$

o bien:

$$(S/N)_i = 40.7 \text{ dB}$$

De nuevo, la razón de señal sobre ruido del sistema será reducida por el ruido del amplificador posdetector, de modo que puede esperarse una $(S/N)_o$ global del sistema de 37 dB aproximadamente. La contribución exacta del ruido y el comportamiento final del

sistema dependen, una vez más, del esquema de modulación/detección, y no es posible asignarlos aquí. Como regla empírica, para una señal digital, una razón de señal sobre ruido aproximada de 20 dB típicamente dará por resultado una proporción de errores en los bits menor de 10^{-9} , que suele ser una meta de funcionamiento generalmente aceptada. Por tanto, este enlace debe tener un margen aproximado de 17 dB.

REFERENCIAS BIBLIOGRÁFICAS

1. N. S. Kapany, *Fiber Optics, Principles and Applications*, Academic Press, Nueva York.
2. N. S. Kapany, "Fiber Optics. V. Light Leakage Due to Frustrated Total Reflection", *J. Opt. Soc. Am.* **49**:770-778; "Fiber Optics. VI. Image Quality and Optical Insulation", 779-787.
3. N. S. Kapany, "An Introduction to Fiber Optics"; N. S. Kapany y colaboradores, "Fiber Optics-Image Transfer on Static and Dynamic Scanning with Fiber Bundles", *J. Opt. Am.* **47**:117.
4. M. B. Panish, "Heterostructure Injection Lasers", *Proc. IEEE* **64**:1512-1540.

5. K. C. Kao y G. A. Hockham, "Dielectric-Fibre Surface Waveguides for Optical Frequencies", *Proc. IEEE* **113**:1151-1158.
6. K. C. y T. W. Davies, "Spectrophotometric Studies of Ultra Low Optical Glasses. I: Single Beam Method", *J. Sci. Instr.* **1** (Series 2):1063-1068.
7. M. W. Jones y K. C. Kao, "Spectrophotometric Studies of Ultra Low Loss Optical Glasses. II: Double Beam Method", *J. Sci. Instr.* **2** (Series 2): 331-335.
8. R. D. Maurer, "Glass Fibers Optical Communications", *Proc. IEEE* **61**:452-462.
9. S. E. Miller y L. C. Tillotson, "Optical Transmission Fiber Guide", *Appl. Opt.* **5**:1538-1549.
10. T. Uchida y colaboradores, "A Light Focusing Fiber Guide", *IEEE J. Quant. Elec.* **QE-5**:331.
11. S. Kawakami y J. Nishizawa, "An Optical Waveguide with the Optimum Distribution of Refractive Index with Reference to Waveform Distortion", *IEEE Trans. MTT* **MTT-16**: 814-818.
12. F. P. Kapron y colaboradores, "Radiation Losses in Glass Optical Waveguides", *Appl. Phys. Lett.* **17**:423-425.
13. S. Shimada, "Systems Engineering for Long Haul Optical-Fiber Transmission", *Proc. IEEE* **68**: 1304-1309.
14. S. E. Miller y colaboradores, "Research Toward Optical Fiber Transmission Systems", *Proc. IEEE* **61**:1703-1751.
15. S. E. Miller y A. G. Chynoweth, eds., *Optical Fiber Telecommunications*, Academic Press, Nueva York.
16. T. Miya y cols., "Ultimate Low-Loss Single-Mode Fibre at 1.55 μm ", *Elec. Lett.* **16**:106-108.
17. D. Marcuse, "Pulse Distortion in Single Mode Fibers", *Appl. Opt.* **19**:1653-1660.
18. H. Tsuchiya y N. Imoto, "Dispersion-Free Single-Mode Fibre in 1.5 μm Wavelength Region", *Elec. Lett.* **15**:476-478.
19. K. Okamoto y colaboradores, "Dispersion Minimization in Single Mode Fibres over a Wide Spectral Range", *Elec. Lett.* **15**:729-731.
20. T. Li, "Structures, Parameters and Transmission Properties of Optical Fibers", *Proc. IEEE* **68**:1175-1179.
21. K. J. Beales y colaboradores, "Multicomponent Glass Fibers for Optical Communications", *Proc. IEEE* **68**:1191-1194.
22. P. C. Schultz, "Fabrication of Optical Waveguides by the Outside Vapor Deposition Process", *Proc. IEEE* **68**:1187-1190.
23. J. B. MacChesney, "Materials and Processes for Preform Fabrication—Modified Chemical Vapor Deposition and Plasma Chemical Vapor Deposition", *Proc. IEEE* **68**:1181-1184.
24. J. G. J. Peelen, *Technical Digest of 4th European Conf. Optical Fiber Communication*, Génova, Italia **37**.
25. T. Izawa y N. Inagaki, "Materials and Processes for Fiber Preform Fabrication—Vapor-Phase Axial Deposition", *Proc. IEEE* **68**:1184-1187.
26. S. E. Miller y colaboradores, "Research Toward Optical Fiber Transmission Systems: Part II, Devices and Systems Considerations", *Proc. IEEE* **61**:1726-1751.
27. R. J. Nelson y colaboradores, "High Output Power in GaAsP ($\lambda = 1.3 \mu\text{m}$) Strip-Buried Heterostructure Lasers", *Appl. Phys. Lett.* **36**:358-360.
28. M. B. Panish y colaboradores, "Double Heterostructure Injection Lasers with Room Temperature Threshold as Low as 2300 A/cm²", *Appl. Phys. Lett.* **16**:326-327.
29. I. Hayashi y colaboradores, "GaAs-Al_xGa_{1-x}As Double Heterostructure Injection Lasers", *J. Appl. Phys.* **42**:1929-1941.
30. B. I. Miller y colaboradores, "Reproducible Liquid-Phase-Epitaxial Growth of Double Heterostructure GaAs-Al_xGa_{1-x}As Laser Diodes", *J. Appl. Phys.* **43**:2817-2826.
31. C. A. Burrus y B. I. Miller, "Small Area, Double Heterostructure Aluminum-Gallium Arsenide Electroluminescent Diode Sources for Optical Fiber Transmission Lines", *Opt. Comm.* **4**:307-309.
32. T. Yamaoka y colaboradores, "GaAlAs LEDs for Fiber-Optical Communication Systems", *Fujitsu Sci. Tech. J.* **14**:133-146.
33. T. P. Lee y colaboradores, "A Stripe Geometry Double-Heterostructure Amplified Spontaneous-Emission (Superluminescent) Diode", *IEEE J. Q. E.* **QE-9**:820-828.
34. Zh. I. Alferov y colaboradores, "High Efficiency Heterojunction LED with Spherical Radiating Surface", *Sov. Tech. Phys. Lett.* **3**:293-294.
35. R. G. Smith, "Photodetectors for Fiber Transmission Systems", *Proc. IEEE* **68**:1247-1253.
36. H. W. Rugg, "An Optimized Avalanche Photodiode", *IEEE Trans. Electron. Dev.* **ED-14**: 239-251.
37. H. Kogelnick, "An Introduction to Integrated Optics", *IEEE Trans. Microwave Theory Tech.* **MTT-23**:2-16.
38. R. V. Schmidt y H. Kogelnik, "Electro-Optically Switched Coupler with Stepped $\Delta\beta$ Reversal Using Ti Diffused LiNbO₃ Waveguides", *Appl. Phys. Lett.* **28**:503-506.
39. R. V. Schmidt y L. L. Buhl, "Experimental 4 \times 4 Optical Switching Network", *Elec. Lett.* **12**: 575-577.
40. M. Papuchon y colaboradores, "Electrically Switched Optical Directional Coupler: Cobra", *Appl. Phys. Lett.* **27**:289-291.
41. R. A. Steinberg y colaboradores, "Polarization-Insensitive Integrated Optical Switches: A New Electrode Design", *Appl. Opt.* **16**:2166-2170.
42. R. A. Soref y colaboradores, "Multimode Achromatic Electro-Optic Waveguide Switch for Fiber-Optic Communications", *Appl. Phys. Lett.* **28**:716-718.
43. W. J. Tomlinson y G. D. Aumiller, "Optical Multiplexer for Multimode Fiber Transmission Systems", *Appl. Phys. Lett.* **31**:169-171.

44. W. J. Tomlinson, "Wavelength Multiplexing in Multimode Optical Fibers", *Appl. Opt.* **16**:2180-2194.
45. R. Watanabe y colaboradores, "Optical Demultiplexer Using Concave Grating in 0.7–0.9 μm Wavelength Region", *Elec. Lett.* **16**:106-108.
46. K. Kobayashi y M. Seki, "Micro-Optic Grating Multiplexers and Optical Isolators for Fiber-Optic Communications", *IEEE J. Quantum Electron.* **QE-16**:11-22.
47. W. J. Tomlinson, "Applications of GRIN-Rod Lenses in Optical Fiber Communication Systems", *Appl. Opt.* **19**:1127-1138.
48. "OD-8601-8604 Optical Directional Coupler", Nippon Electric Co. Ltd., Data Sheet DPR-047E.
49. A. H. Fitch y colaboradores, "Four Channel Fiber Optic Link Employing Wavelength Division Multiplexing and Full Duplex Transmission", Topical Meeting on Optical Fiber Communications, Phoenix, AZ.
50. B. S. Kawasaki y K. O. Hill, "Low Loss Access Coupler for Multimode Optical Fiber Distribution Networks", *Appl. Opt.* **16**:1794-1795.
51. J. F. Dalgleish, "Splices, Connectors and Power Couplers for Field and Office Use", *Proc. IEEE* **68**:1226-1232.

48.5 TÉCNICAS AVANZADAS PARA LAS COMUNICACIONES DIGITALES*

Bernard Sklar

48.5.1 Introducción

Durante las dos décadas pasadas ha surgido una variedad impresionante de técnicas de procesamiento de señales de comunicaciones. En esta sección se presenta un panorama general de algunas de estas técnicas, especialmente la forma en que se relacionan con las comunicaciones digitales vía satélite. La información se expone en el contexto de una estructura empleada para seguir las etapas de procesamiento. Las transformaciones están organizadas por clases funcionales: formateo y codificación en la fuente, modulación, codificación en el canal, multicanalización y acceso múltiple, dispersión de frecuencia, cifrado y sincronización.

La comunicación por satélite presenta dos características únicas: la capacidad de cobertura de la Tierra, con una flexibilidad que no es posible reproducir con enlaces terrestres, y la disponibilidad de ancho de banda, que excede cualquier otro ancho de banda previamente disponible para las comunicaciones intercontinentales¹. La mayor parte de las comunicaciones vía satélite hasta la fecha son de naturaleza analógica. Sin

embargo, la comunicación digital se está volviendo cada vez más atractiva debido a la siempre creciente demanda de la comunicación de datos y debido a que la transmisión digital ofrece opciones de procesamiento de datos y flexibilidades que no se tienen con la transmisión analógica².

En esta sección se presenta un panorama general de las comunicaciones digitales; sin embargo, el tratamiento casi siempre es en el contexto de un enlace de comunicaciones vía satélite. La característica clave de un sistema de comunicación digital (DCS, de *digital communication system*) es que envía solamente un conjunto finito de datos, en contraste con un sistema de comunicación analógico, que es capaz de enviar un conjunto infinito de mensajes. En un DCS, el objetivo en el receptor no es reproducir con precisión una onda de forma, sino determinar con base en una señal perturbada con ruido cuál o cuáles ondas de forma del conjunto finito se enviaron por medio del transmisor. Una medida importante del rendimiento del sistema es el número promedio de decisiones erróneas que se efectúan, o la probabilidad de error (P_E).

En la figura 48-111 se ilustra un DCS típico. Sean M símbolos o mensajes m_1, m_2, \dots, m_M por transmitir, y representémoslos cada símbolo por una onda de forma correspondiente $s_1(t), s_2(t), \dots, s_M(t)$. El símbolo (o mensaje) m_i se envía mediante la transmisión de la onda de forma digital $s_i(t)$ durante T segundos, que es el período del símbolo. El siguiente símbolo se envía durante el siguiente período. Dado que M símbolos pueden representarse con $k = \log_2 M$ dígitos binarios (bits), la velocidad de transmisión de datos puede expresarse como:

$$R = (1/T)\log_2 M = k/T \text{ bits por segundo (b/s)}$$

La velocidad de transmisión de datos suele expresarse en bps, se trate o no de dígitos binarios. Un símbolo binario es el caso especial caracterizado por $M = 2$ y $k = 1$. Se considera que una onda de forma digital significa una onda de forma de un voltaje o corriente que representa un símbolo digital. La onda de forma está dotada con características especialmente elegidas de amplitud, frecuencia o fase, que permiten la selección de una onda de forma diferente para cada símbolo de un conjunto finito de símbolos. A diferentes puntos a lo largo de la trayectoria de la señal el ruido corrompe las ondas de forma $s(t)$, de modo que su recepción debe denominarse estimación $\hat{s}(t)$. Tal ruido y su efecto dañino para el rendimiento del sistema se analizan en la sección 48.5.3.

48.5.2 Etapas de procesamiento de la señal

El diagrama funcional de bloques de la figura 48-111 ilustra el flujo de datos a través del DCS. Los bloques superiores, que se identifican como formateo, codificación en la fuente, cifrado, codificación en el canal, multicanalización, modulación, dispersión de frecuencia y acceso múltiple, determinan las transformaciones de la señal desde la fuente hasta el transmisor. Los bloques inferiores determinan las transformaciones de

*Una versión de esta sección apareció como una serie en dos partes en el IEEE Communications Magazine bajo el título "A Structured Overview of Digital Communications—A Tutorial Review". Dicho artículo ganó el Communications Society Prize Paper Award para el Communications Magazine.

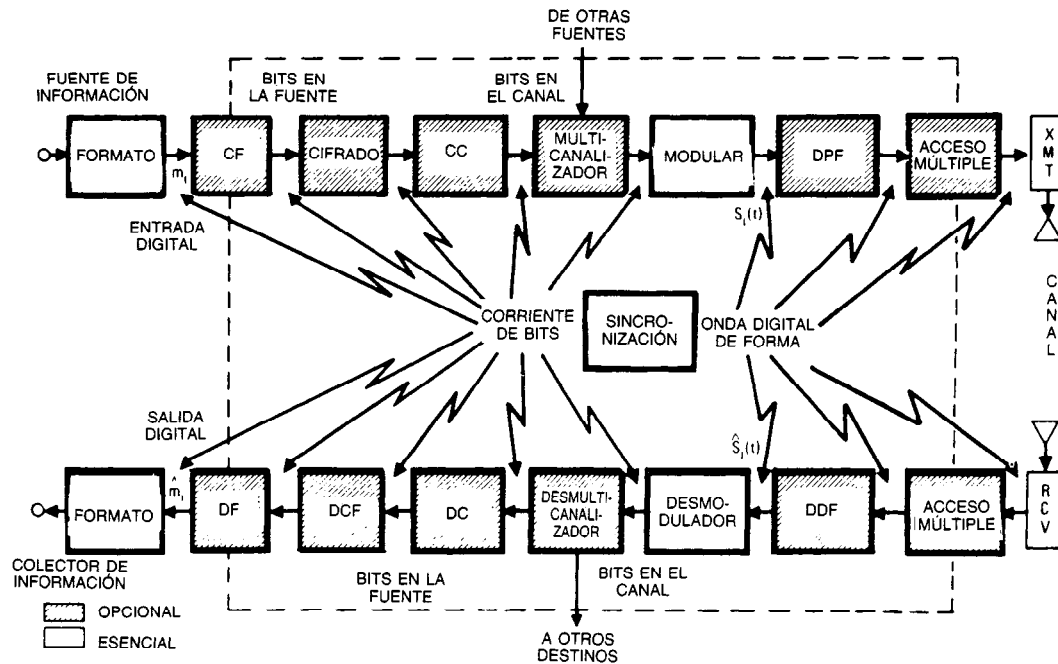


Fig. 48-111. Diagrama de bloques de un sistema típico de comunicaciones digitales. CF = codificación en la fuente; CC = codificación en el canal; DPF = dispersión de frecuencia; DF = decodificación en la fuente; DCF = descifrado; DC = decodificación en el canal; DDF = desdispersión de frecuencia.

la señal desde el receptor de regreso a la fuente; esencialmente los mismos invierten las etapas del procesamiento de la señal efectuadas por los bloques superiores. Los bloques, que se encuentran encerrados por la línea discontinua, inicialmente sólo consistían en las funciones del modulador y del desmodulador; de ahí el nombre "modem". Durante las dos décadas pasadas, con frecuencia se incorporaron otras funciones para el procesamiento de la señal dentro del mismo ensamblaje del modulador y desmodulador. Por consiguiente, el término "modem" a menudo abarca los pasos de procesamiento enmarcados por la línea discontinua de la figura 48-111. Cuando éste es el caso, el modem puede concebirse como el "cerebro" del sistema, y el transmisor y el receptor como los "músculos". Mientras que el transmisor consta de una etapa de conversión ascendente de frecuencia, un amplificador de alta potencia y una antena, la parte del receptor está ocupada por una antena, un amplificador de entrada (preamplificador) de bajo ruido y una etapa de conversión descendente de frecuencia, por lo general a una frecuencia intermedia (IF).

De todas las etapas de procesamiento de la señal, solamente el formateo, modulación y desmodulación son esenciales para todos los DCS; las otras etapas de procesamiento dentro del modem se consideran opciones de diseño para diversas necesidades del sistema. La codificación en la fuente, como aquí se define, elimina redundancias de información y efectúa la conversión de analógico a digital (A/D). El cifrado evita que usuarios no autorizados comprendan los mensajes y que se

introduzcan mensajes falsos en el sistema. La codificación en el canal puede, para una velocidad de transmisión de datos dada, mejorar el rendimiento P_E a expensas de potencia o ancho de banda, reducir el requisito del ancho de banda del sistema a expensas de potencia o rendimiento P_E , o bien reducir el requisito de potencia a expensas del ancho de banda o rendimiento P_E . La distribución de frecuencias hace que la señal sea menos vulnerable a interferencias (tanto naturales como intencionales), y puede utilizarse para proporcionar privacidad a los comunicadores. La multicanalización y el acceso múltiple combinan señales que podrían tener características diferentes o provenir de fuentes diferentes.

El flujo de las etapas para el procesamiento de la señal que se muestra en la figura 48-111 representa una disposición típica; sin embargo, algunas veces los bloques se implementan en un orden distinto. Por ejemplo, la multicanalización puede efectuarse antes que la codificación del canal, que la modulación o —con un proceso de modulación en dos etapas (subportadora y portadora)— puede realizarse entre dos etapas. De forma parecida, la distribución puede efectuarse en cualquier punto a lo largo de la cadena de transmisión; su localización exacta depende de la técnica específica que se utilice. En la figura 48-111 se ilustra el aspecto recíproco del procedimiento; cualesquiera etapas del procesamiento de la señal que se realicen en la cadena de transmisión deben invertirse en la cadena de recepción. En la figura también se indica que, desde la fuente hasta el modulador, un mensaje asume la forma de

una corriente de bits, también denominada señal de banda base. Después de la modulación, el mensaje asume la forma de un senoide codificado digitalmente (onda de forma digital). De modo similar, un mensaje recibido se presenta como onda de forma digital hasta que es desmodulado. A partir de entonces asume la forma de una corriente de bits para todas las etapas posteriores del procesamiento de la señal.

En la figura 48-112 se muestran las funciones básicas del procesamiento de señal, que pueden considerarse transformaciones desde el espacio de una señal hasta el de otra. Las transformaciones están clasificadas en siete grupos fundamentales: formateo y codificación en la fuente, modulación, codificación en el canal, multicanalización y acceso múltiple, dispersión, cifrado y sincronización. La organización presenta alguna superposición inherente; a pesar de ello, proporciona una estructura útil para este panorama general. El texto de Lindsey y Simon³ es una excelente referencia para las transformaciones de modulación, codificación y sincronización tratadas aquí. Los extensos libros de Spilker⁴ y de Bhargava y colaboradores⁵ tratan de forma específica las comunicaciones digitales vía satélite. A continuación se considerará por separado cada una de las siete transformaciones básicas por orden general de importancia.

Formateo y codificación en la fuente

El primer paso esencial en el procesamiento, el formateo, hace que los datos comunicados sean compatibles

para el procesamiento digital. El formateo se define como cualquier operación que transforme datos en símbolos digitales. La codificación en la fuente significa la compresión y adición de datos para el formateo. Algunos autores consideran el formateo un caso especial de codificación en la fuente (para lo cual la compresión de los datos es cero), en vez de establecer una diferencia entre ambos. La fuente de casi todos los datos comunicados (excepto en el caso de las transformaciones de computadora a computadora que ya se encuentran en forma digital) es de naturaleza textual o analógica. Si los datos consisten en un texto alfanumérico, éste se codifica por caracteres con alguno de los varios formatos estándares, tales como el American Standard Code for Information Interchange (ASCII), el Extended Binary Coded Decimal Interchange Code (EBCDIC), o el Baudot, con lo que se lleva a forma digital. Si los datos son analógicos, primero es necesario muestrear la onda de forma (limitada en banda) a una velocidad de por lo menos $2f_m$ Hz (la frecuencia de Nyquist), donde f_m es la mayor frecuencia contenida en la onda de forma. Tal muestreo asegura la reconstrucción perfecta de la señal analógica; el submuestreo da por resultado un fenómeno denominado *aliasing* (aparición de una frecuencia más baja), que introduce errores. Sin embargo, la velocidad mínima de muestreo puede ser menor de $2f_m$ si la frecuencia más baja de la señal contenida en la onda de forma es diferente de cero⁶. La cuantización de las muestras temporales permite que cada muestra sea expresada como un

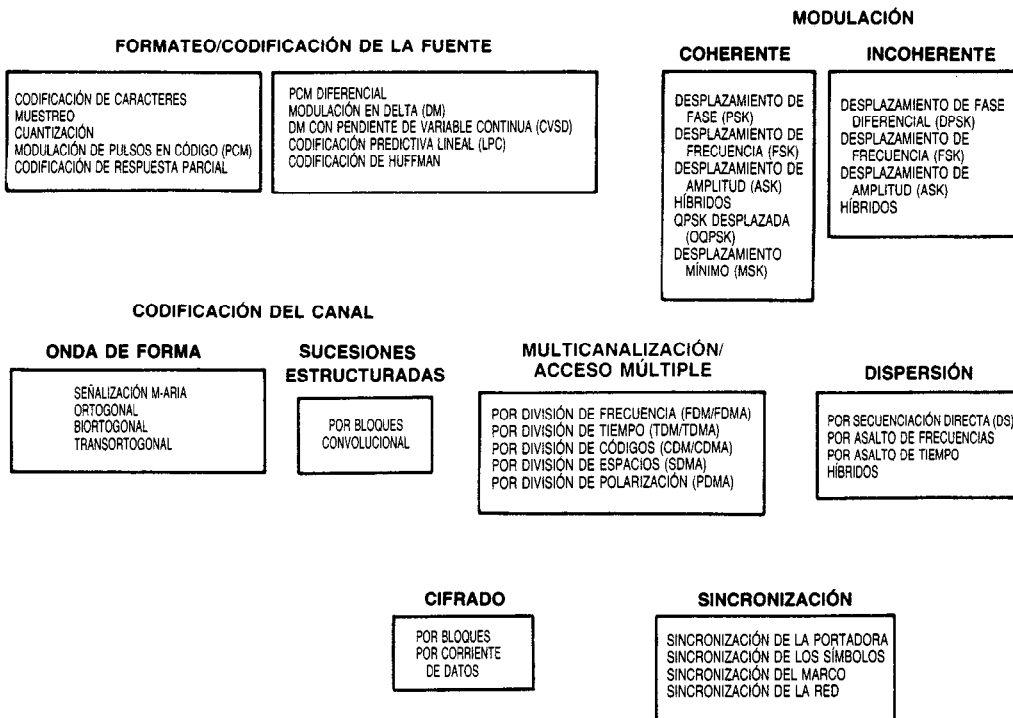


Fig. 48-112. Transformaciones básicas de comunicación digital.

nivel entre un número finito de niveles predeterminados; cada uno de estos niveles puede representarse por medio de un símbolo digital. Después de la cuantización todavía es posible recuperar la onda de forma analógica, aunque no de manera precisa; es posible mejorar la fidelidad de la reconstrucción de la onda de forma analógica incrementando el número de niveles de cuantización (para lo cual se requiere un aumento en el ancho de banda de transmisión).

La modulación de pulsos en código o modulación por codificación de pulsos (PCM, de *pulse code modulation*), que es el formato digital clásico y de mayor uso, convierte las muestras cuantizadas en grupos codificados de pulsos de dos niveles en los que se emplean amplitudes fijas. Cada grupo de pulsos representa el valor de una amplitud cuantizada expresado en notación binaria. Existen varios subformatos PCM (como el sin regreso a cero, el Manchester y el Miller), cada uno de los cuales tiene alguna característica especial, tal como el autocronometraje o una identificación espectral compacta³. La codificación duobinaria o de respuesta parcial (también denominada codificación correlativa) es una técnica de formateo con la que se mejora la eficiencia del ancho de banda mediante la introducción de una interferencia controlada entre los símbolos. La técnica también proporciona capacidades para la detección de errores sin la introducción de redundancias en la corriente de datos⁷⁻⁹.

Tanto la codificación en la fuente como el formateo significan la codificación de datos en la fuente con un formato digital (conversión A/D); sólo en este sentido los dos son idénticos. Sin embargo, el término "codificación en la fuente" ha adquirido otros significados adicionales en el empleo de los DCS. Además del formateo digital, la "codificación en la fuente" también denota la compresión de datos (o reducción en la velocidad de transmisión de datos). Con la conversión A/D estándar empleando PCM, la compresión de datos solamente puede lograrse reduciendo la velocidad de muestreo o el número de niveles de cuantización por muestra, procesos ambos que incrementan el error cuadrático medio de la señal reconstruida. Las técnicas de codificación en la fuente realizan la reducción de la velocidad mediante la eliminación de la redundancia presente en la mayor parte de las transmisiones de mensajes, sin sacrificar la fidelidad de reconstrucción. Se dice que una fuente de datos digitales presenta redundancia si los símbolos no son igualmente probables o si no son estadísticamente independientes. La codificación en la fuente puede reducir la velocidad de transmisión de datos si existe cualquiera de estas condiciones. A continuación se proporcionan algunas descripciones de las más frecuentes técnicas de codificación en la fuente.

En la PCM diferencial (DPCM) se utilizan las diferencias entre muestras, en vez de su amplitud verdadera. Para la mayor parte de los datos, la variación media de amplitud de muestra es menor que la variación total de amplitud; por consiguiente, se requieren menos bits para describir la diferencia. Los sistemas DPCM en realidad codifican la diferencia entre una muestra de amplitud presente y un valor de amplitud

predicho, estimado con base en muestras pasadas. El decodificador utiliza un algoritmo parecido para decodificar. La modulación en delta (DM, de *delta modulation*) es el nombre otorgado al caso especial de DPCM en el que el nivel de cuantización de la salida se toma como igual a 1 bit. Aunque la DM pueda implementarse con facilidad, adolece de "sobrecarga de pendiente", condición en la cual la pendiente de la señal de llegada excede la capacidad del sistema para seguir de cerca la fuente analógica a la velocidad de muestreo dada. A fin de mejorar el rendimiento, siempre que se detecte una sobrecarga de pendiente puede hacerse variar la ganancia del sistema conforme a un algoritmo predeterminado conocido por el receptor. Si el sistema se diseña para hacer variar adaptativamente la ganancia en un intervalo continuo, la modulación se denomina modulación en delta con pendiente de variable continua (CVSD, de *continuous variable slope delta*), o modulación en delta adaptativa (ADM, de *adaptive delta modulation*). Con la modulación CVSD se ha obtenido una codificación del habla de calidad aceptable a velocidades de transferencia de bits menores de 25 kb/s, que es una notable reducción en la velocidad de los datos cuando se le compara con los 56 kb/s de la PCM utilizada en sistemas telefónicos comerciales¹⁰.

Otro ejemplo de codificación en la fuente es la codificación predictiva lineal (LPC, de *linear predictive coding*). Esta técnica es de utilidad cuando la onda de forma resulta de un proceso que puede modelarse como un sistema lineal. En vez de codificar muestras de la onda de forma que esté codificándose, se codifican las características importantes del proceso. Para el habla, entre tales características se incluyen la ganancia, separación e información con o sin habla. Mientras que en la PCM cada muestra se procesa en forma independiente, en un sistema predictivo como la DPCM se utiliza una suma ponderada de las n muestras pasadas a fin de predecir cada muestra actual; después se transmite la señal del "error". Los pesos se calculan a fin de minimizar la energía promedio en la señal de error que representa la diferencia entre las amplitudes predicha y verdadera. Para el habla, los pesos se calculan sobre cortos segmentos de la onda de forma, de 10 a 30 ms, por lo que cambian a medida que varían los estadísticos del habla. La técnica LPC se ha utilizado para producir una calidad aceptable del habla a una velocidad de transmisión de datos de 2.4 kb/s y alta calidad a 7.2 kb/s¹¹⁻¹³. Las perspectivas actuales del formateo digital del habla pueden consultarse en la obra de Crochiere y Flanagan¹⁴.

En algunas técnicas de codificación en la fuente se utilizan sucesiones de códigos de longitud desigual para minimizar el número promedio de bits requeridos por muestra de datos. Para la compresión de datos en cualquier conjunto de símbolos puede utilizarse un procedimiento útil de codificación, denominado codificación de Huffman^{15,16}, con la condición de que se conozca la probabilidad *a priori* de ocurrencia de los símbolos y de que éstos no sean equiprobables. La codificación de Huffman genera una sucesión binaria para cada símbolo a fin de obtener el menor número promedio de bits por muestra para las probabilidades a

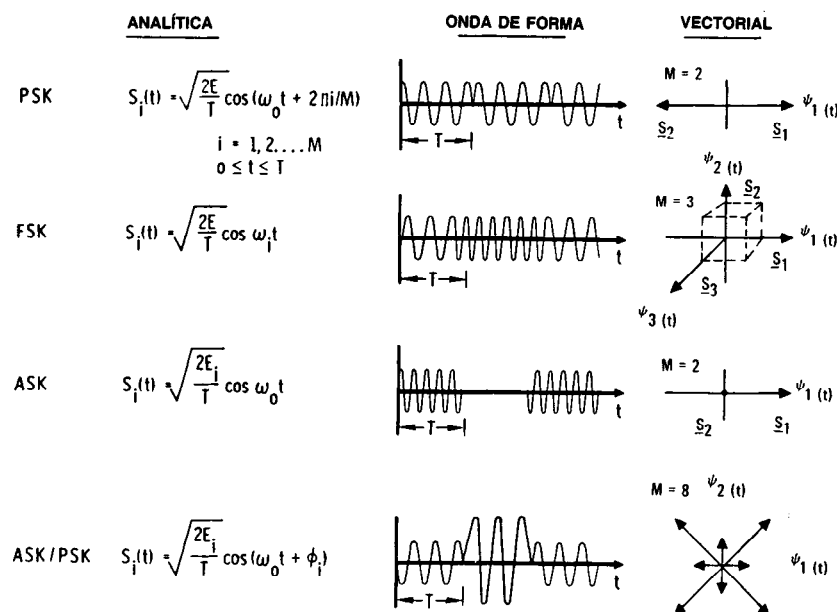


Fig. 48-113. Formatos de modulación digital. PSK = manipulación por desplazamiento de fase; FSK = manipulación por desplazamiento de frecuencia; ASK = manipulación por desplazamiento de amplitud; ASK/PSK = combinación híbrida.

priori dadas. La técnica implica la asignación de sucesiones de código más cortas a los símbolos de mayor probabilidad, y sucesiones de código mayores a aquellos con menor probabilidad. El precio que debe pagarse para lograr de esta forma la reducción de la velocidad de transmisión de los datos es un incremento proporcional en la complejidad del decodificador. Además, existe la tendencia de que los símbolos erróneos se propaguen.

Técnicas de modulación digital

Formatos de modulación digital. En términos generales, la modulación es un proceso mediante el cual se hace variar alguna característica de una onda de forma conforme a otra onda de forma. Un sinusoide tiene exactamente tres características que es posible utilizar para distinguirlo de otros sinusoides: fase, frecuencia y amplitud. Para el objeto de la radiotransmisión, la modulación se define como el proceso por medio del cual se hacen variar la fase, la frecuencia o la amplitud de una onda portadora de radiofrecuencia (RF) de conformidad con la información por transmitir. En la figura 48-113 se ilustran ejemplos de formatos de modulación digital: manipulación por desplazamiento de fase o defasamiento (PSK, de *phase shift keying*), manipulación por desplazamiento de frecuencias (FSK, de *frequency shift keying*), manipulación por desplazamiento de la amplitud (ASK, de *amplitude shift keying*) y una combinación híbrida de ASK y PSK, que algunas veces se denomina modulación de amplitud en cuadratura (QAM, de *quadrature amplitude modulation*). En la primera columna se observa la expresión

analítica, en la segunda se presenta un esquema de la onda de forma y en la tercera se muestra una representación vectorial. En el caso general de señalización M -aria, el procesador acepta k bits de la fuente a la vez e instruye al modulador para producir una onda de forma entre un conjunto disponible de $M = 2^k$ tipos de ondas de forma. La modulación binaria, donde $k = 1$, es sólo un caso especial de la modulación M -aria. Para el ejemplo binario PSK (BPSK) de la figura 48-113, M es igual a dos tipos de ondas de forma (binario). Para el ejemplo FSK, M es igual a tres tipos de ondas de forma (terciario); obsérvese que este valor $M = 3$ para la FSK se ha elegido a fin de destacar los ejes perpendiculares entre sí. En la práctica, M suele ser igual a una potencia de 2 que no sea cero (2, 4, 8, 16, ...). Para el ejemplo ASK, M es igual a dos tipos de ondas de forma; para el ejemplo ASK/PSK, M es igual a ocho tipos de ondas de forma (8-ario). La representación vectorial para cada tipo de modulación (excepto la FSK) se caracteriza sobre un plano cuyas coordenadas polares representen la amplitud y fase de la señal. A los conjuntos de señales que es posible representar como vectores opuestos (con diferencias de fase iguales a 180°) en tal plano, por ejemplo los BPSK, se les denomina señales antípodas. En el caso de la modulación FSK, la representación vectorial se caracteriza por medio de coordenadas cartesianas tales que cada uno de los ejes mutuamente perpendiculares representa una frecuencia de transmisión diferente. A los conjuntos de señales que es posible caracterizar con tales ejes ortogonales se les llama señales ortogonales. En la sección 48.1 se analizan con más detalle las técnicas de modulación.

La **modulación** se definió como el proceso mediante el cual se hace variar una portadora o subportadora mediante una señal de la banda base; en la figura 48-112 se presenta la jerarquía de la modulación digital. Cuando en el receptor se aprovecha el conocimiento de la referencia de fase de la onda portadora para detectar las señales, el proceso se denomina **detección coherente**; cuando el receptor no cuenta con información de referencia de fase, el proceso se denomina **incoherente**. En la detección coherente ideal, en el receptor existen a disposición prototipos de las posibles señales de llegada. Estas ondas de forma prototípicas reproducen exactamente en todos los aspectos el conjunto de señales, incluso en cuanto a fase RF. Se dice entonces que el receptor está acoplado en fase con el transmisor. Durante la detección, el receptor multiplica e integra (correlaciona) la señal de llegada con cada una de sus réplicas prototípicas. Bajo el encabezado de modulación coherente (fig. 48-112), la PSK, la FSK y la ASK también se presentan como combinaciones híbridas.

La modulación incoherente se refiere a sistemas diseñados para operar sin conocimiento de la fase; no se requiere el procesamiento para la estimación de fase, de modo que una ventaja sobre los sistemas coherentes es la reducción de complejidad, y la modulación intermedia es una mayor P_E . En la figura 48-112 se muestra que los tipos de modulación enumerados en la columna de modulación incoherente son casi los mismos que se observan en la columna correspondiente a la modulación coherente. La única diferencia es que no puede haber "PSK incoherente", debido a que incoherente significa sin empleo de la información de fase. Sin embargo, existe una técnica "pseudoPSK" denominada PSK diferencial (DPSK) en la que se utiliza la información de fase RF del símbolo precedente como referencia de fase para detectar el símbolo presente (lo que se describe más adelante, en la desmodulación).

Dos esquemas de modulación digital, de interés especial para ser utilizados en canales no lineales limitados en banda, son la PSK cuadrifásica escalonada (o desplazada) (SQPSK, de *staggered quadrature PSK*), o bien OQPSK (de *offset quadrature PSK*) y la manipulación por desplazamiento mínimo (MSK, de *minimum shift keying*). En ambas técnicas se retienen bajos niveles espectrales de los lóbulos laterales, a la vez que se permite un aceptable funcionamiento para la detección. La generación de ambas técnicas puede representarse como dos sistemas binarios ortogonales antípodas, en los que los símbolos en los dos canales están defasados por la mitad de la duración de un símbolo. En la OQPSK se utilizan pulsos de formas rectangulares, y en la MSK, sinusoides de medio ciclo. Debido a la configuración sinusoidal de los pulsos en la MSK, dicha técnica puede considerarse una FSK de fase continua con desviación de frecuencia igual a la mitad de la velocidad de transmisión de los bits^{17,18}.

Desmodulación. El análisis de todos los esquemas de desmodulación o detección coherentes implica el concepto de distancia entre una onda de forma recibida desconocida y un conjunto de ondas de forma conoci-

das. En un espacio de señales descrito por medio de ejes mutuamente perpendiculares es fácil definir mediciones de distancia de tipo euclidiano. Puede demostrarse¹⁹ que cualquier conjunto finito arbitrario de ondas de forma $s_i(t)$, donde $s_i(t)$ es físicamente realizable y tiene duración T , puede expresarse como una combinación lineal de N ondas de forma ortonormales $\phi_1(t)$, $\phi_2(t)$, ..., $\phi_N(t)$, tal que:

$$s_i(t) = \sum_{j=1}^N a_{ij} \phi_j(t) \quad (48-130)$$

donde:

$$a_{ij} = \int_0^T s_i(t) \phi_j(t) dt \quad i = 1, 2, \dots, M \quad 0 \leq t \leq T$$

$$j = 1, 2, \dots, N \quad (48-131)$$

$$N \leq M$$

y:

$$\int_0^T \phi_i(t) \phi_j(t) dt = 1 \quad \text{para } i = j$$

$$= 0 \quad \text{en caso contrario} \quad (48-132)$$

El ruido gaussiano blanco aditivo (AWGN, de *additive white Gaussian noise*) puede expresarse como una combinación lineal de las ondas de forma ortonormales:

$$n(t) = \sum_{j=1}^N n_j \phi_j(t) + \tilde{n}(t) \quad (48-133)$$

donde:

$$n_j = \int_0^T n(t) \phi_j(t) dt$$

Para el problema de la detección de señales, es posible separar el ruido en dos componentes:

$$n(t) = \hat{n}(t) + \tilde{n}(t) \quad (48-134)$$

donde:

$$\hat{n}(t) = \sum_{j=1}^N n_j \phi_j(t)$$

se toma como el ruido dentro del espacio de las señales, o como la proyección de las componentes del ruido sobre los ejes de la señal $\phi_1(t)$, $\phi_2(t)$, ..., $\phi_N(t)$, y:

$$\tilde{n}(t) = n(t) - \hat{n}(t)$$

se define como el ruido fuera del espacio de las señales. En otras palabras, $\tilde{n}(t)$ podría concebirse como el ruido que es sintonizado de manera efectiva por el detector. El símbolo $\hat{n}(t)$ representa el ruido que interferirá con el proceso de detección; en lo sucesivo se denominará sencillamente $n(t)$.

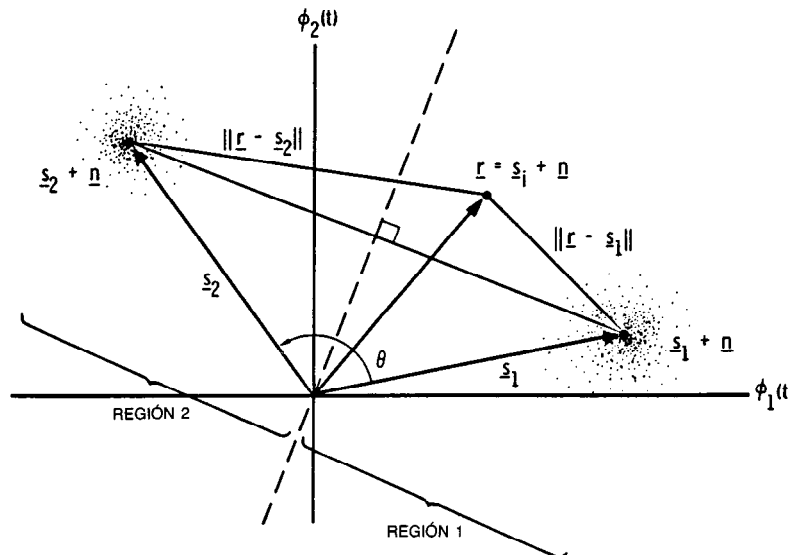


Fig. 48-114. Espacio bidimensional de señales.

Una vez que se ha adoptado un conjunto idóneo de N funciones ortonormales [obsérvese que $\phi(t)$ no se ha restringido a ninguna forma específica], cada una de las ondas de forma de la señal transmitida $s_i(t)$ está completamente determinada por el vector de sus coeficientes:

$$s_i = (a_{i1}, a_{i2}, \dots, a_{iN}) \quad i = 1, 2, \dots, M$$

De forma parecida, el ruido $n(t)$ puede expresarse mediante el vector de sus coeficientes:

$$\mathbf{n} = (n_1, n_2, \dots, n_N)$$

donde \mathbf{n} es un vector aleatorio con media cero y distribución gaussiana.

Debido a que *cualquier* conjunto arbitrario de ondas de forma, así como el ruido, puede representarse como una combinación lineal de ondas de forma ortonormales, es posible utilizar una distancia de tipo euclidiano en tal espacio ortonormal como criterio de decisión para la detección de *cualquier* conjunto de señales en presencia de AWGN.

Detección en presencia de AWGN. En la figura 48-114 se ilustra un espacio de señales bidimensional, el lugar geométrico de dos prototipos de señales binarias perturbadas por ruido ($s_1 + \mathbf{n}$) y ($s_2 + \mathbf{n}$), y una señal recibida \mathbf{r} . En notación vectorial, la señal recibida es $\mathbf{r} = s_i + \mathbf{n}$, donde $i = 1$ o 2 . Esta representación geométrica o vectorial de las señales y del ruido facilita el análisis de la detección de la señal digital. Los vectores s_1 y s_2 son fijos, ya que las ondas de forma $s_1(t)$ y $s_2(t)$ no son aleatorias. El vector o punto \mathbf{n} es un vector aleatorio; por tanto, también \mathbf{r} es un vector aleatorio.

La tarea del detector después de recibir a \mathbf{r} es decidir cuál de las señales s_1 o s_2 fue realmente transmitida. El método suele ser decidir la clasificación de la señal que

produzca la P_E mínima, aunque también son posibles otras estrategias²⁰. Para el caso en que M sea igual a dos clases de señal, con s_1 y s_2 igualmente probables y con el ruido AWGN, resulta que la regla para decidir con un error mínimo es: Siempre que la señal recibida \mathbf{r} caiga en la región 1, elegir la señal s_1 ; cuando caiga en la región 2, elegir la señal s_2 (fig. 48-114). Una afirmación equivalente es: elegir la clase de señal tal que la distancia $d(\mathbf{r}, s) = \|\mathbf{r} - s_i\|$ sea mínima, donde $\|\mathbf{x}\|$ se denomina "norma" del vector \mathbf{x} y es una generalización del concepto de longitud.

Detección de PSK coherente. En la figura 48-115 se ilustra la estructura del receptor implicado por la regla anterior. Existe un integrador de producto (correlacionador) para cada onda de forma prototípica (M en todas); los correlacionadores son seguidos por una etapa de decisión. La señal recibida está correlacionada con cada onda de forma prototípica conocida *a priori* por el receptor. En la etapa de decisión se elige la señal perteneciente al correlacionador con mayor salida (la mayor z_i). Por ejemplo, sea:

$$s_1(t) = \sin \omega t$$

$$s_2(t) = -\sin \omega t$$

$$n(t) = \text{proceso aleatorio con media cero y distribución gaussiana}$$

Supóngase que se transmitió $s_1(t)$, de modo que:

$$r(t) = s_1(t) + n(t) \quad \text{y} \quad z_i = \int_0^T r(t)s_i(t)dt \quad i = 1, 2$$

Los valores esperados de los integradores de producto, como se ilustra en la figura 48-115, se encuentran de la siguiente manera:

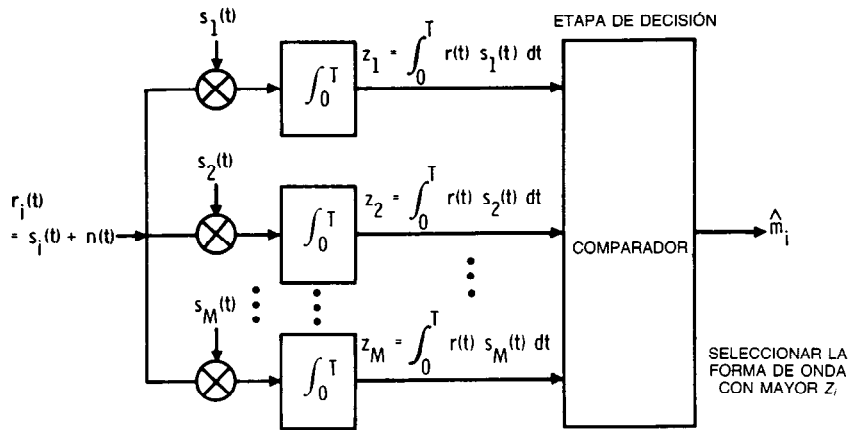


Fig. 48-115. Integradores de producto (correlacionadoras) y etapa de decisión.

$$E[z_1(t = T)] = E\left[\int_0^T \sin^2 \omega t + n(t) \sin \omega t dt\right] = T/2$$

$$E[z_2(t = T)] = E\left[\int_0^T -\sin^2 \omega t + n(t) \sin \omega t dt\right] = -T/2$$

donde E es el promedio estadístico.

Midiendo la localización de las señales en el espacio de señales, la etapa de decisión debe decidir cuál de ellas fue transmitida. La regla de decisión es elegir la señal con el mayor valor de z_i . A menos que el ruido sea alto y de naturaleza capaz de producir algún error, se considera que la señal recibida es $s_1(t)$. En presencia de ruido este proceso es estadístico; el detector óptimo

es aquel que en promedio comete menos errores. La única estrategia que puede utilizar el detector es "conjeturar", utilizando alguna regla de decisión optimizada.

En la figura 48-116 se muestra el proceso de detección teniendo presente el espacio de las señales. Representa un sistema coherente PSK de cuatro niveles (4-ario) o un sistema QPSK. En los términos utilizados antes para la señalización M -aria, $k = 2$ y $M = 2^2 = 4$. Los dígitos de la fuente binaria se colectan de dos en dos (dos cada vez), y para cada intervalo de símbolos los dos dígitos consecutivos instruyen al modulador acerca de cuál de las cuatro ondas de forma producir. En general, para sistemas PSK (MPSK) M -arios coherentes, $s_i(t)$ puede expresarse como:

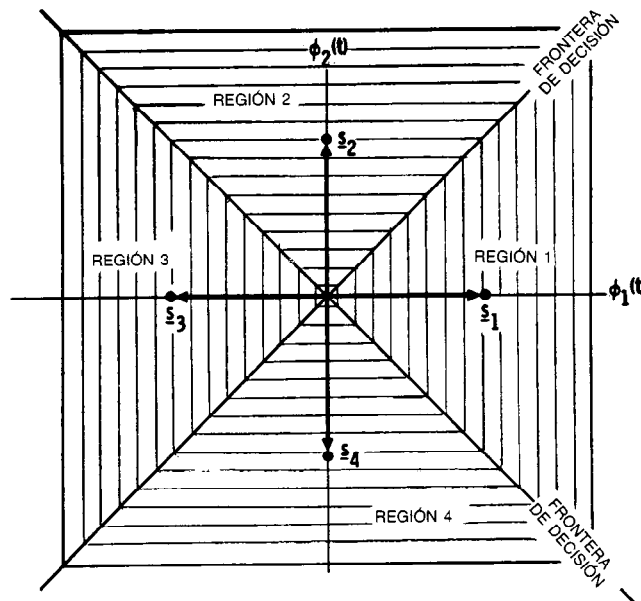


Fig. 48-116. Espacio de señales y regiones de decisión para la detección mediante manipulación por desplazamiento de fase en cuadratura (QPSK).

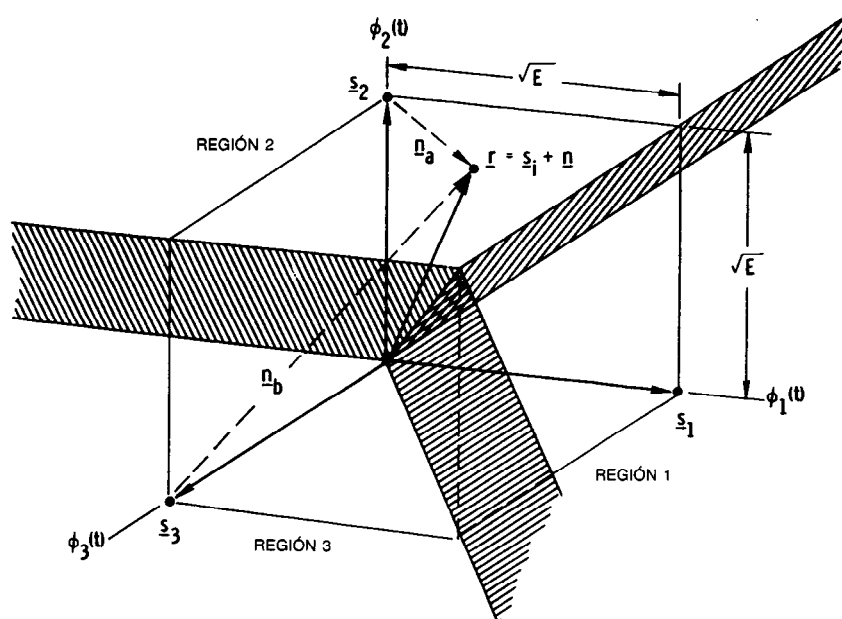


Fig. 48-117. Espacio de señales y regiones de decisión para la detección mediante manipulación por desplazamiento de frecuencia (FSK) coherente terciaria.

$$s_i(t) = \sqrt{E/T} \cos(\omega_0 t - 2\pi i/M) \quad \text{para } 0 \leq t \leq T \quad s_i(t) = \sqrt{2E/T} \cos \omega_i t \quad \text{para } 0 \leq t \leq T \quad i = 1, 2, \dots, M$$

donde $i = 1, 2, \dots, M$ $= 0$ en caso contrario

Aquí, E es el contenido de energía de $s_i(t)$ y ω_0 es un múltiplo entero de $2\pi/T$.

Puede elegirse un conjunto adecuado de ejes ortogonales escalados de modo que se satisfaga la ecuación 48-132 como sigue:

$$\begin{aligned} \phi_1(t) &= \sqrt{2/T} \cos \omega_0 t \\ \phi_2(t) &= \sqrt{2/T} \sin \omega_0 t \end{aligned} \quad (48-135)$$

Ahora es posible expresar $s_i(t)$ en términos de estas coordenadas ortogonales, de donde se obtiene:

$$s_i(t) = \sqrt{E} \cos(2\pi i/M) \phi_1(t) + \sqrt{E} \sin(2\pi i/M) \phi_2(t) \quad (48-136)$$

La regla de decisión para el detector (fig. 48-116) es decidir que $s_1(t)$ fue transmitida si el punto de la señal recibida cae en la región 1, que $s_2(t)$ fue transmitida si el punto de la señal recibida cae en la región 2, etc. En otras palabras, la regla de decisión es elegir la i -ésima onda de forma con el mayor valor de la salida del correlacionador z_i (fig. 48-115).

Detección de FSK coherente. La modulación FSK se caracteriza por la información contenida en la frecuencia de la onda portadora. Un conjunto típico de ondas de forma de la señal se describe por medio de las expresiones:

donde E es el contenido de energía de $s_i(t)$ y $(\omega_{i+1} - \omega_i)$ es un múltiplo entero de $2\pi/T$. La forma de mayor utilidad para las coordenadas ortonormales $\phi_1(t), \phi_2(t), \dots, \phi_N(t)$ es la que se indica a continuación:

$$\phi_j(t) = \sqrt{2/T} \cos \omega_j t \quad j = 1, 2, \dots, N$$

y, con base en la ecuación 48-131:

$$a_{ij} = \int_0^T \sqrt{2E/T} \cos \omega_i t \sqrt{2/T} \cos \omega_j t dt$$

Por consiguiente:

$$\begin{aligned} a_{ij} &= \sqrt{E} \quad \text{para } i = j \\ &= 0 \quad \text{en caso contrario} \end{aligned}$$

En otras palabras, el i -ésimo punto de la señal está localizado en el i -ésimo eje coordenado a una distancia \sqrt{E} del origen del espacio de señales. En la figura 48-117 se ilustran los vectores (puntos) de la señal y las regiones de decisión para una modulación FSK coherente terciaria ($M = 3$). En este esquema, la distancia entre dos puntos cualesquiera de la señal, s_i y s_j , es constante:

$$d(s_i, s_j) = \|s_i - s_j\| = \sqrt{2E} \quad \text{para } i \neq j$$

Como en el caso de la PSK coherente, el espacio de señales se divide en M regiones distintas, cada una con un punto de la señal prototípica. La regla óptima de decisión es decidir que la señal transmitida pertenece a la clase cuyo número de índice es igual a la región en que se encontró la señal recibida. En la figura 48-117 se muestra un punto \mathbf{r} de la señal recibida en la región 2. Con la regla de decisión, el detector la clasifica como señal \mathbf{s}_2 . Debido a que el ruido es un vector aleatorio, existe una probabilidad mayor que cero de que la localización de \mathbf{r} se deba a alguna otra señal diferente de \mathbf{s}_2 . Por ejemplo, si el transmisor envió \mathbf{s}_2 , entonces \mathbf{r} es la suma de $\mathbf{s}_2 + \mathbf{n}_a$ y por tanto la decisión de elegir \mathbf{s}_2 es correcta; sin embargo, si el transmisor envió en realidad \mathbf{s}_3 , entonces \mathbf{r} debe ser la suma de $\mathbf{s}_3 + \mathbf{n}_b$ (fig. 48-117), y la decisión de elegir \mathbf{s}_2 es errónea.

Detección de DPSK. Con los sistemas incoherentes, no se toman medidas para sincronizar en fase el receptor con el transmisor. Así, si la onda transmitida es:

$$s_i(t) = \sqrt{2E/T} \cos(\omega_0 t + \phi_i) \quad i = 1, 2, \dots, M$$

entonces la señal recibida puede caracterizarse como:

$$r(t) = \sqrt{2E/T} \cos(\omega_0 t + \phi_i + \alpha) + n(t)$$

donde α es una incógnita que se supone es una variable aleatoria uniformemente distribuida entre cero y 2π .

Para la detección coherente se utilizan los integradores de producto (o sus equivalentes); para la detección incoherente tal uso suele ser inadecuado, debido a que la salida de un integrador de producto es función del ángulo desconocido α . Sin embargo, si se supone que α varía tan lentamente que es posible considerarlo constante durante dos periodos de tiempo ($2T$), entonces la diferencia de fase relativa entre dos ondas de forma consecutivas es independiente de α ; es decir:

$$(\phi_1 + \alpha) - (\phi_2 + \alpha) = \phi_1 - \phi_2$$

Ésta es la base de la modulación DPSK. La fase de la portadora del intervalo precedente de señalización se utiliza como referencia de fase para la desmodulación. Su empleo requiere la codificación diferencial de la

sucesión del mensaje en el transmisor, ya que la información es transportada por la diferencia de fase entre dos ondas de forma consecutivas. Para evitar el i -ésimo mensaje ($i = 1, 2, \dots, M$), la onda de forma de la señal actual debe tener su fase adelantada $2\pi i/M$ radianes respecto a la de la onda de forma precedente. Entonces el detector es capaz de calcular las coordenadas de la señal que esté llegando, integrándola en producto con las ondas de forma generadas localmente $\sqrt{2/T} \cos \omega_0 t$ y $\sqrt{2/T} \sin \omega_0 t$. De esta manera, el detector mide el ángulo entre los puntos actual y previamente recibido de la señal (fig. 48-118)¹⁹.

Una forma de concebir la diferencia entre la PSK coherente y la DPSK es que en la primera la señal recibida se compara con una referencia limpia; sin embargo, en la segunda se comparan entre sí dos señales con ruido. Por consiguiente, es posible afirmar que en la DPSK hay dos veces más ruido que en la PSK. Así, cuando se compara con la PSK, la DSPK manifiesta una degradación de 3 dB aproximadamente; esta cantidad decrece rápidamente con una razón creciente de señal sobre ruido. En general, los errores tienden a propagarse (en periodos de tiempo adyacentes) debido a la correlación entre las ondas de forma de la señalización. La solución para esta pérdida de rendimiento es un descenso en la complejidad del sistema.

Detección de FSK incoherente. Un detector FSK incoherente puede implementarse con correlacionadores (fig. 48-115). Sin embargo, el hardware debe configurarse como un detector de energía, sin aprovechar las mediciones de fase. Es por ello por lo que se implementa con doble número de ramificaciones de canal que el detector coherente. En la figura 48-119 se ilustran los canales en fase (I) y los canales en cuadratura (Q) utilizados para detectar incoherentemente el conjunto de señales. En otra implementación posible se utilizan filtros seguidos de detectores de envolvente; los detectores están acoplados a las envolventes de las señales y no a las señales en sí. La fase de la portadora carece de importancia en la definición de la envolvente, por lo que no se utiliza información de fase. En el caso de la FSK, la decisión de si fue transmitido un 1 o un 0 se toma con base en cuál de los dos detectores de envolvente presenta la mayor amplitud en el mo-

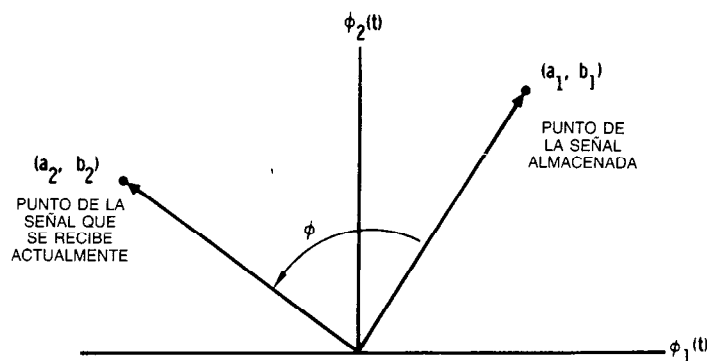


Fig. 48-118. Espacio de señales para la detección mediante manipulación por desplazamiento diferencial de fase (DPSK).

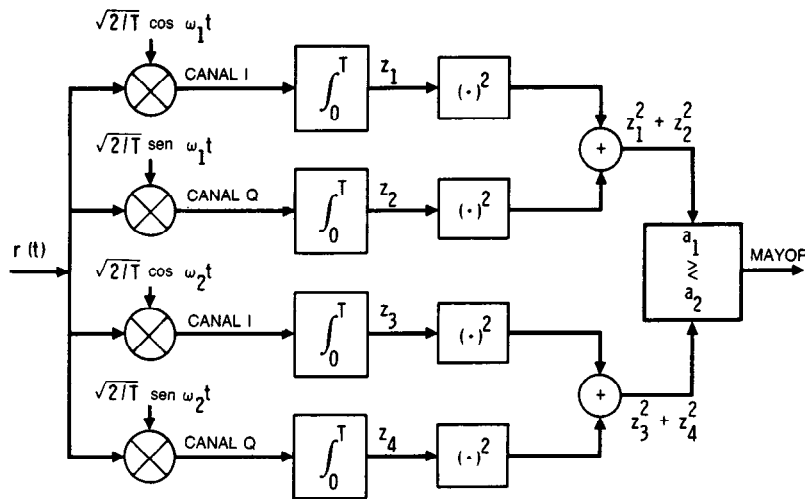


Fig. 48-119. Receptor en cuadratura para la detección FSK incoherente.

mento de la medición. De forma similar, para un sistema de manipulación por desplazamiento de frecuencias múltiples (FSK M -aria, o MFSK), la decisión acerca de cuál de las M señales fue transmitida se toma con base en cuál de los M detectores de envolvente tiene la máxima salida.

Probabilidad de error. Los cálculos de la probabilidad de error (P_E), que puede observarse geoméricamente en la figura 48-114, implican el cálculo de la probabilidad de que dada una señal particular, por ejemplo s_1 , el vector de ruido \mathbf{n} produzca una señal recibida que caiga fuera de la región 1; éste es el objetivo de todos los cálculos de P_E . Para el caso general de señalización M -aria, la probabilidad de tomar una decisión incorrecta se denomina probabilidad de error en los símbolos, o sencillamente P_E . A menudo es conveniente especificar el rendimiento de los sistemas por medio de la probabilidad de error en los bits (P_E), incluso cuando las decisiones se tomen con base en los símbolos para los cuales $k > 1$.

Para señales ortogonales, P_E y P_B están relacionadas como sigue²¹:

$$P_B/P_E = (2^{k-1})/(2^k - 1)$$

Para esquemas no ortogonales, tales como la señalización MPSK, suele utilizarse una codificación de binario a M -ario tal que las sucesiones binarias correspondientes a símbolos adyacentes (defasamientos) difieran sólo en una posición de bit; uno de tales códigos es el Gray. Cuando ocurre un error M -ario en los símbolos, es más probable que solamente uno de los k bits de entrada esté equivocado. Para tales señales³:

$$P_B \cong P_E/\log_2 M = P_E/k \quad \text{para } P_E \ll 1$$

Por conveniencia, el análisis está restringido a la modulación BPSK ($k = 1$, $M = 2$). Para el caso binario, la probabilidad de error en los símbolos es igual a la

probabilidad de error en los bits. Supóngase que se ha transmitido la señal $s_1(t)$ y que $r(t) = s_1(t) + n(t)$. Con la hipótesis de señales equiprobables, y recordando que la decisión de la región 1 contra la región 2 depende de los integradores de producto y de la etapa de decisión (fig. 48-115), puede expresarse:

$$P_{E|\text{binario}} = P_B = \Pr \left[\int_0^T r(t)s_2(t) dt > \int_0^T r(t)s_1(t) dt \mid r(t) = s_1(t) + n(t) \right]$$

para $0 \leq t \leq T$. La solución para la expresión P_B es, en el caso del ruido gaussiano blanco aditivo (sec. 47.1):

$$P_B = 1/\sqrt{2\pi} \int_{\sqrt{E_b/N_0(1-\cos\theta)}}^{\infty} \exp(-u^2/2) du$$

donde E_b es la energía de la señal por bit en joules, N_0 es la densidad de ruido en el receptor en watts por hertz, y θ es el ángulo entre s_1 y s_2 (fig. 48-114). Cuando $\theta = \pi$, las señales se denominan antípodas y P_B se transforma en:

$$P_B = 1/\sqrt{2\pi} \int_{\sqrt{2E_b/N_0}}^{\infty} \exp(-u^2/2) du \equiv Q(\sqrt{2E_b/N_0}) \quad (48-137)$$

La misma clase de análisis se aplica a fin de determinar las expresiones de P_B para los otros tipos de modulación. El parámetro E_b/N_0 de la ecuación 48-137 puede expresarse como la razón de potencia promedio de la señal entre potencia promedio del ruido, S/N (o SNR). Por medio de la introducción arbitraria del ancho de banda de la señal de la banda base W , es posible expresar las siguientes identidades, que demuestran la relación entre E_b/N_0 y SNR:

$$E_b/N_0 = ST/N_0 = S/RN_0 = SW/RN_0W = (S/N)(W/R) \quad (48-138)$$

donde S = potencia promedio de la señal de modulación

T = duración de un bit

$R = 1/T$ = velocidad de transmisión de los bits

$N = N_0W$

La razón adimensional E_b/N_0 (necesaria para lograr una P_B especificada) se utiliza uniformemente para caracterizar el rendimiento del sistema de comunicaciones digitales. Obsérvese que la detección óptima de la señal digital implica la implementación de un correlador (o filtro doble), en cuyo caso el ancho de banda de la señal es igual al ancho de banda del ruido. A menudo es necesario trabajar con un modelo de sistema para el cual éste no es el caso (subóptimo); en la práctica, en el parámetro requerido E_b/N_0 solamente se refleja un factor que explique el rendimiento subóptimo de la detección. Por tanto, el parámetro E_b/N_0 requerido puede considerarse una métrica que caracteriza el rendimiento de un sistema en comparación con el rendimiento del otro; cuanto menor sea el E_b/N_0 requerido, más eficiente serán la modulación del sistema y el proceso de detección.

Las expresiones de P_B para los esquemas de modulación binarios analizados se enumeran en la tabla 48-19 y se comparan en la figura 48-120. Para SNR grandes puede observarse que existe una diferencia aproximada de 4 dB entre la mejor (PSK coherente) y la peor (FSK incoherente). Algunas veces 4 dB es un precio bajo para la sencillez de la implementación ganada al ir de la PSK coherente a la FSK incoherente; sin embargo, para algunas aplicaciones 1 dB de ahorro sigue valiendo la pena. Además de la P_B y la complejidad del sistema, existen otras consideraciones; por ejemplo, en algunos casos (tales como en condiciones de propagación con desvanecimiento aleatorio), un sistema incoherente es más resistente y deseable, ya que puede haber dificultades con una referencia coherente.

Tabla 48-19. Probabilidad de errores en los bits (P_B) para esquemas selectos de modulación binaria

Modulación	P_B^a
PSK coherente	$Q\left[\sqrt{\frac{2E_b}{N_0}}\right]$
DPSK incoherente	$1/2 \exp(-E_b/N_0)$
FSK coherente	$Q\left[\sqrt{\frac{E_b}{N_0}}\right]$
FSK incoherente	$1/2 \exp[-1/2(E_b/N_0)]$

^a Aquí E_b/N_0 = (energía/bit)/densidad de ruido = S/N_0R = potencia de la señal recibida/(densidad del ruido \times velocidad de transferencia de los bits) y $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \exp(-u^2/2) du$; $Q(x) \approx (1/x\sqrt{2\pi}) \exp(-x^2/2)$, para $x > 3$.

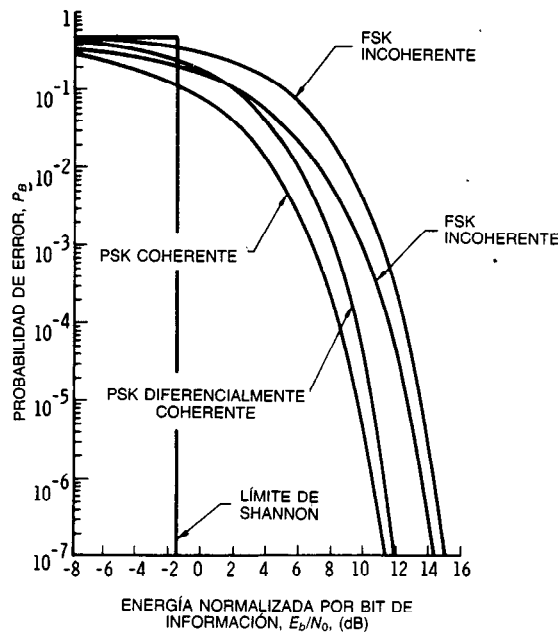


Fig. 48-120. Probabilidad de errores en los bits para esquemas de modulación binaria selectos. FSK = manipulación por desplazamiento de frecuencia; PSK = manipulación por desplazamiento de fase.

Merece la pena mencionar una excepción a la tabla 48-19 y a la figura 48-120, a la luz de los actuales esquemas eficientes de modulación del ancho de banda. La modulación MSK, que puede considerarse coherente, manifiesta un rendimiento de proporción de errores igual al de la BPSK cuando se detecta con un receptor adecuado¹⁸.

Codificación en el canal

La codificación en el canal (figs. 48-111 y 48-112) se refiere a la transformación de datos, efectuada después de la codificación en la fuente, pero antes de la modulación, en la que se transforman bits de la fuente en bits del canal. La codificación en el canal se divide en dos grupos: la codificación de la onda de forma y en sucesiones estructuradas (fig. 48-112).

Con base en esto, la **codificación de la onda de forma** (o diseño de la señal) se definirá como cualquier transformación de datos que haga que el proceso de detección sea menos susceptible a errores y que por ello mejore el rendimiento de la transmisión. Puede considerarse mejor como una transformación que demuestre un rendimiento mejorado en un sentido global o "estructurado" (gestalt), debido a que la codificación produce un conjunto de señales con mejores propiedades de distancia que las del conjunto original de señales. La categoría de sucesiones estructuradas, en comparación, mejora el rendimiento al fijar los datos con redundancia estructurada, que después puede utilizarse para detectar y corregir errores de transmisión.

Codificación de la onda de forma. La señalización M -aria se describió como un esquema de modulación o codificación que procesa k bits a la vez. El sistema dirige al modulador para que elija una de sus $M = 2^k$ ondas de forma para cada sucesión de k bits, donde M es el tamaño del conjunto de símbolos y k es el número de dígitos binarios que representa cada símbolo. La señalización M -aria sola, para el caso en que $k > 1$, puede considerarse un procedimiento de codificación de la onda de forma que afecta el rendimiento del sistema. Las señales ortogonales manifiestan una P_B mejorada a expensas del ancho de banda a medida que k aumenta; las señales no ortogonales manifiestan eficiencia de ancho de banda mejorada (R/W) a expensas de la potencia o del rendimiento de P_B . R/W se mide en bits por segundo por hertz (b/s/Hz); para la señalización binaria, el valor típico de la eficiencia del ancho de banda es aproximadamente 1 b/s/Hz. Sin embargo, los sistemas actuales de manipulación por desplazamiento de fases múltiples (MPSK) y de modulación de amplitud en cuadratura (QAM) a menudo presentan eficiencias de ancho de banda de 3 b/s/Hz o más^{22,23}.

Otro ejemplo de codificación de la onda de forma es el empleo de un conjunto mejorado de señales en sustitución de los símbolos originales de los datos. Los más conocidos de estos códigos se denominan conjuntos ortogonales y biortogonales de señales. Se dice que el conjunto ortogonal de señales $s_i(t)$, donde $i = 1, 2, \dots, M$, es ortogonal si y sólo si:

$$z_{ij} = 1/E \int_0^T s_i(t)s_j(t)dt = 1 \quad \text{para } i = j \\ = 0 \quad \text{en caso contrario} \quad (48-139)$$

donde se supone que todas las M señales tienen igual energía E , y que T es la duración del símbolo. Así como la señalización M -aria con formato de modulación ortogonal (tal como la MFSK) mejora el rendimiento en P_B con k creciente, un conjunto de señales construido ortogonalmente, antes de la modulación MPSK, produce la misma mejora.

Un conjunto biortogonal de señales puede obtenerse a partir de un conjunto ortogonal de $M/2$ señales aumentando éste con el negativo de cada señal. El conjunto biortogonal consiste en una combinación de señales ortogonales y antípodas. Debido a que los vectores de las señales antípodas tienen mejores propiedades de distancia que las ortogonales (fig. 48-113), los códigos biortogonales se comportan ligeramente mejor que los ortogonales. Con respecto a z_{ij} de la ecuación 48-139, los códigos biortogonales pueden caracterizarse como:

$$z_{ij} = 1 \quad \text{para } i = j \\ = -1 \quad \text{para } i \neq j, |i - j| = M/2 \\ = 0 \quad \text{para } i \neq j, |i - j| \neq M/2$$

Un código generado a partir de un conjunto ortogonal al borrar el primer dígito de cada palabra de código se denomina código transortogonal o símplex. Tal código

representa la mínima energía equivalente (en el sentido de P_B) del igualmente probable conjunto ortogonal. Al comparar el rendimiento de los códigos ortogonal, biortogonal y símplex, es posible afirmar que el código símplex requiere la mínima SNR para una proporción específica de errores en los bits. Sin embargo, para k grande los tres esquemas son esencialmente idénticos en cuanto al rendimiento a medida que tienden al límite de Shannon de -1.6 dB^{5,24}. La codificación biortogonal requiere la mitad de ancho de banda que las demás. A pesar de ello, cada uno de los requisitos de ancho de banda de los códigos (y la complejidad del sistema) crece exponencialmente con el valor de k .

Sucesiones estructuradas (códigos lineales de bloques).

La codificación en el canal con sucesiones estructuradas representa un método para insertar redundancia estructurada en la fuente de datos, de modo que los errores de transmisión puedan identificarse. Las sucesiones estructuradas pueden dividirse en dos subcategorías importantes: para codificación por bloques y para codificación convolucional (fig. 48-112). Con la codificación por bloques, primero se divide la fuente de datos en bloques de k bits de datos cada uno; cada bloque puede representar uno de $M = 2^k$ mensajes diferentes. El codificador transforma cada bloque de mensajes en un bloque más grande de n dígitos. Este conjunto de 2^k mensajes cifrados se denomina bloque codificado. Los $(n - k)$ dígitos que el codificador agrega a cada bloque de mensajes se denominan dígitos redundantes; no transportan nueva información. La razón de bits de datos entre el número total de bits en el bloque, k/n , se denomina proporción de código. El código en sí se denomina código (n, k) .

Supóngase que $\mathbf{c} = (c_1, c_2, \dots, c_n)$, donde $c_j = 1$ o 0 , es la palabra de código transmitida, de un conjunto $i = 1, \dots, M$ palabras cifradas, y que $\mathbf{r} = (r_1, r_2, \dots, r_n)$ es la sucesión recibida en la entrada del codificador. Si todas las palabras de código tienen la misma probabilidad de ser transmitidas, el esquema óptimo de decodificación por utilizar se denomina decodificación de máxima probabilidad; es parecido al esquema óptimo de desmodulación con hipótesis *a priori* semejantes. El decodificador calcula la probabilidad condicional $P(\mathbf{r} | \mathbf{c}_i)$ para todas las 2^k palabras cifradas. La palabra cifrada \mathbf{c}_i se identifica como la palabra transmitida si $P(\mathbf{r} | \mathbf{c}_i)$ es la máxima de las probabilidades calculadas.

La posible mejora en rendimiento, con la codificación en el canal, puede ilustrarse con el siguiente ejemplo de un código de corrección de un solo error (15,11). La notación (15,11) significa que cada bloque de 15 bits comprende 11 bits de datos y 4 bits de redundancia. Considérese la siguiente transmisión sin codificar. Supóngase modulación BPSK: $S/N_0 = 43$ 776; velocidad de transferencia de datos $R = 4$ 800 b/s. P_B^U y P_M^U representan las probabilidades codificadas de errores en los bits y en el mensaje, respectivamente.

Sin codificación. Con base en la ecuación 48-138:

$$E_b/N_0 = S/RN_0 = 9.12 \quad (=9.6 \text{ dB})$$

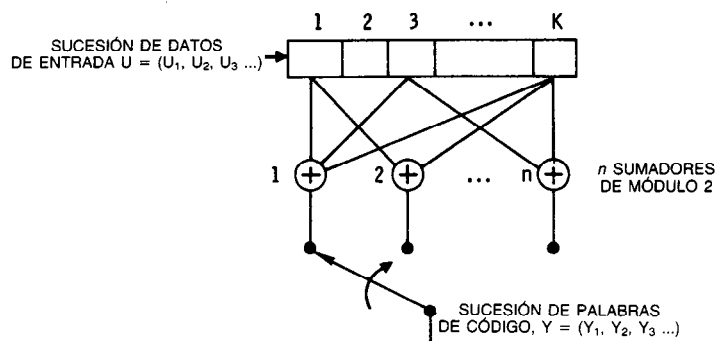


Fig. 48-121. Codificador convolucional con longitud K restringida y velocidad $1/n$.

$$P_B^U = Q(\sqrt{2E_b/N_0}) = Q(\sqrt{18.24}) = 1.02 \times 10^{-5} \quad (48-140)$$

$$P_B^C \cong \binom{15}{2} (P_B^C)^2 (1 - P_B^C)^{13} = 1.94 \times 10^{-6} \quad (48-143)$$

donde:

$$Q(x) \cong (1/x\sqrt{2\pi}) \exp(-x^2/2) \text{ para } x > 3$$

$$P_M^U = 1 - (1 - P_M^U)^{11} = 1.12 \times 10^{-4} \quad (48-141)$$

Con codificación. La velocidad de transferencia de bits codificados R_C es 15/11 veces la velocidad de transferencia de bits de datos:

$$R_C = 4\,800 \times 15/11 \cong 6\,545 \text{ b/s}$$

$$E_b/N_0 = S/R_C N_0 = 6.688 \text{ (-8.25 dB)}$$

El E_b/N_0 para el bit codificado es algo menor que para el bit no codificado, debido a que la velocidad de transferencia de bits ha aumentado, pero se supone que la potencia del transmisor es fija:

$$P_B^C = Q(\sqrt{2E_b/N_0}) = Q(\sqrt{13.38}) = 1.36 \times 10^{-4} \quad (48-142)$$

Por comparación con las ecuaciones 48-140 y 48-142 puede observarse que la proporción de errores en los bits se ha degradado; debe ser posible detectar más bits durante el mismo intervalo y con la misma potencia disponible. La mejora de rendimiento debida a la codificación todavía no es evidente. Ahora se calcula la proporción de errores en el mensaje codificado P_M^C como:

$$P_M^C = \sum_{k=2}^{15} \binom{15}{k} (P_B^C)^k (1 - P_B^C)^{15-k}$$

La sumatoria se inicia con $k = 2$, ya que el código corrige todos los errores únicos dentro de un bloque de $n = 15$ bits. Una buena aproximación se obtiene utilizando solamente el primer término de la sumatoria. Para P_B^C , se utiliza el valor calculado en la ecuación 48-142.

La ecuación 48-143 indica la proporción de errores en el mensaje para el bloque de 15 bits codificados. En este ejemplo típico puede observarse, comparando las ecuaciones 48-141 y 48-143, que la probabilidad de error del mensaje ha mejorado por un factor de 58, gracias al empleo de un código de bloques.

La mayor parte de la investigación acerca de códigos de bloques se ha concentrado en una subclase de códigos lineales, conocida como códigos cíclicos. Una palabra de código cíclico, al cabo de cualquier número de desplazamientos cíclicos, tiene la propiedad de seguir siendo una palabra codificada válida del conjunto original de palabras codificadas. Los códigos cíclicos son atractivos debido a que es fácil implementarlos empleando registros de desplazamiento con retroalimentación. Los métodos de decodificación son sencillos y eficaces, y por lo general se utilizan los mismos registros de desplazamientos con retroalimentación que se emplean para la codificación^{25,26}.

Sucesiones estructurales (codificación convolucional).

Un codificador convolucional convoluciona una sucesión de datos de entrada con una función de codificación (fig. 48-121). El codificador es mecanizado con un registro de desplazamiento de K etapas y n sumadores módulo 2, donde K se denomina restricción de longitud. La fuente se representa por medio de la sucesión de bits de entrada $U = (u_1, u_2, \dots, u_i, \dots)$. En la i -ésima unidad de tiempo el bit de datos u_i es desplazado a la primera etapa; todos los bits presentes en el registro se desplazan una etapa a la derecha, y la salida de los n sumadores se muestrea y transmite secuencialmente. Ya que existen n bits de código por bit de datos, en este caso la velocidad de codificación es $1/n$. Para el caso general, k bits a la vez se desplazan en el registro, y la velocidad de codificación es k/n . Los n bits de código que ocurren al tiempo i constituyen la i -ésima rama de la palabra de código $Y_i = (y_{1i}, y_{2i}, \dots, y_{ni})$. La palabra de código Y consiste en la sucesión de ramas $Y = (Y_1, Y_2, \dots)$.

Sea el estado del codificador en el instante i definido por $X_i = (u_{i-1}, u_{i-2}, \dots, u_{i-k+1})$, que es el contenido de

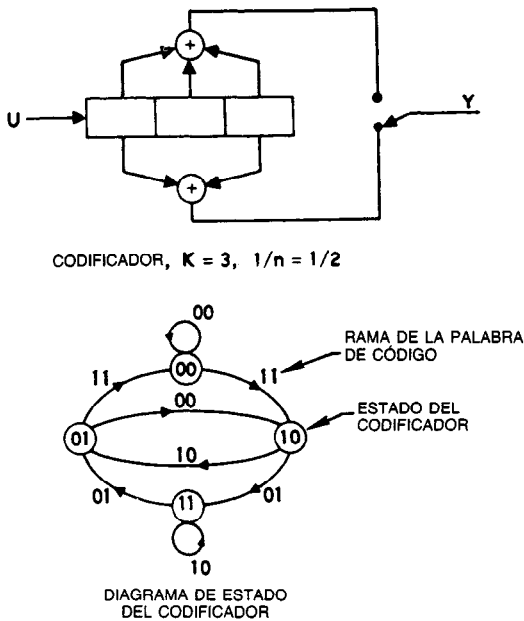


Fig. 48-122. Caracterización del codificador por medio de un diagrama de estado.

las $K - 1$ etapas en el extremo derecho del registro de desplazamiento. La rama Y_i de la i -ésima palabra de código está determinada completamente por la etapa X_i y por el bit de entrada actual u_i ; por tanto, el estado X_i representa los antecedentes del codificador al determinar la salida del codificador. Se dice que el estado del codificador es de Markov, en el sentido de que la probabilidad $P(X_{i+1} | X_i, X_{i-1}, \dots, X_0)$ de encontrarse en el estado X_{i+1} , dados todos los estados previos, depende solamente del estado X_i ; es decir, la probabilidad es igual a $P(X_{i+1} | X_i)$. Una forma sencilla de representar un codificador es por medio de un diagrama de estado, como se muestra en la figura 48-122 para el codificador con $K = 3$. Los estados del diagrama

representan los posibles contenidos de la $K - 1$ etapas del extremo derecho del registro de desplazamiento. Existen dos transiciones a partir de cada estado, correspondientes a los dos posibles bits de entrada, y existe una rama de palabra de código asociada a cada transición. El diagrama de estado también puede usarse para obtener una función de transferencia, que a su vez puede utilizarse con objeto de deducir cotas para la probabilidad de error²⁷.

Una forma conveniente de incorporar la historia temporal del codificador es mediante el diagrama de enrejado que se muestra en la figura 48-123. Para cada unidad de tiempo el enrejado muestra todas las posibles transiciones entre estados. Existen dos trayectorias posibles que parten de cada estado, correspondientes a los dos valores posibles del bit de entrada de datos. Por convención, una línea discontinua en el enrejado corresponde al dato de entrada 1, y una línea continua, al dato de entrada 0. También se muestran las ramas de la palabra de código de salida correspondientes a las transiciones para el codificador de la figura 48-122; se presentan como identificaciones en las ramas del enrejado.

Un decodificador óptimo efectúa la estimación de máxima probabilidad de la palabra de código transmitida Y , dada la observación Z . El decodificador elige la palabra de código \hat{Y} si $P(Z | \hat{Y}) = \max_Y P(Z | Y)$. Debido a que se supone que el ruido es independiente:

$$P(Z|Y) = \prod_{i=1}^{\infty} P(Z_i|Y_i) = \prod_{i=1}^{\infty} \prod_{j=1}^n P(z_{ji}|y_{ji})$$

donde Y_i es la i -ésima rama de la palabra de código Y , Z_i es la i -ésima rama de la sucesión recibida Z , z_{ji} es el j -ésimo bit de código de Z_i , e y_{ji} es el j -ésimo bit de código de Y_i ; cada rama abarca n bits de código. El problema del decodificador consiste en elegir una trayectoria a través del enrejado de la figura 48-123 (cada trayectoria posible define una palabra de código) de modo que $\prod_{i=1}^{\infty} \prod_{j=1}^n P(z_{ji} | y_{ji})$ sea máximo.

Los canales binarios se caracterizan por la necesidad de tomar decisiones inflexibles (decisiones de dos ni-

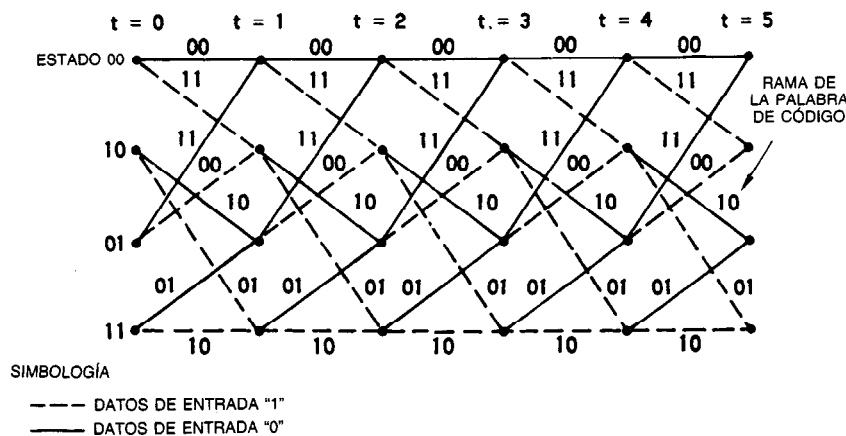


Fig. 48-123. Diagrama de enrejado de codificación.

veles) en los bits de código recibidos. Los canales continuos se caracterizan por la capacidad de tomar decisiones flexibles (decisiones de niveles múltiples). Una decisión de este segundo tipo puede concebirse como una decisión con un factor de confianza asociado. La capacidad de tomar decisiones flexibles a lo largo del proceso de decodificación da por resultado mejor funcionamiento de decodificación (aprox. 2 dB) que si se toman decisiones inflexibles. Por último, para un sistema digital todas las decisiones deberán convertirse en decisiones inflexibles; es decir, 1 o 0.

Un decodificador de máxima probabilidad "de fuerza bruta" calcula la probabilidad de los datos recibidos en todas las trayectorias a través del enrejado. El número de trayectorias para una sucesión de información con L bits es 2^L ; el método de la fuerza bruta se vuelve impráctico a medida que aumenta L . El algoritmo de Viterbi realiza esencialmente la decodificación de máxima probabilidad; sin embargo, reduce la carga computacional al aprovechar la estructura especial en el enrejado del código. La ventaja de la decodificación de Viterbi (en comparación con la de fuerza bruta) es que la complejidad del decodificador es una función lineal, en vez de exponencial, de L^{28} . El algoritmo implica el cálculo de una métrica (medida de similitud) entre la señal recibida (en el instante t_i) y todas las trayectorias del enrejado que entran a cada estado (en el mismo instante t_i), donde $i = 1, 2, \dots$. En caso de que dos trayectorias que terminen en un estado dado sean redundantes, se almacena la que tenga mayor métrica (trayectoria superviviente). Esta selección del superviviente se efectúa para todas las trayectorias que entran a cada uno de los demás estados. El decodificador continúa de esta forma avanzando más profundamente en el enrejado, tomando decisiones por medio de la eliminación de las trayectorias con menor probabilidad. Las trayectorias supervivientes deben ser almacenadas en un intervalo de cinco longitudes de restricción aproximadamente, a fin de considerar el retardo de codificación. Los requisitos de almacenamiento crecen exponencialmente con la longitud restringida; el estado tecnológico actual limita los decodificadores de Viterbi a $K = 10$ aproximadamente. Estos decodificadores son bastante efectivos en cuanto a costo para proporciones de error moderadas, pero no son capaces de lograr eficazmente proporciones muy bajas de error. Por otra parte, son capaces de velocidades muy elevadas, cuando los decodificadores secuenciales se vuelven demasiado costosos²⁹.

Un decodificador secuencial funciona generando hipótesis acerca de la sucesión de datos transmitidos; genera una métrica entre sus hipótesis y la señal recibida. Prosigue hacia delante mientras la métrica permanece dentro de la tolerancia; en caso contrario se regresa, cambiando hipótesis hasta que encuentra una métrica mejorada por medio de una búsqueda por ensayo y error. Es posible implementarlo a fin de trabajar con decisiones inflexibles o flexibles, aunque por lo general estas últimas se evitan debido a que incrementan bastante el almacenamiento y los cálculos requeridos. La complejidad del decodificador es relativamente insensible a la longitud de restricción del código; por ello,

las longitudes de restricción suelen hacerse muy grandes ($K = 40$), lo que es un factor importante a fin de obtener tal rendimiento de P_B baja. El número de hipótesis deficientes y búsquedas hacia atrás son función de la SNR; con mayor ruido deben generarse más hipótesis. Debido a esta variabilidad en la carga computacional es necesario contar con almacenamiento en acumuladores ("buffers") intermedios. Algunas veces estos buffers se sobresaturarán, por lo que dejarán una sección de datos en su forma codificada. Así, una parte importante en una especificación de un decodificador secuencial es la probabilidad de sobresaturación ("desbordamiento") del buffer.

El funcionamiento de estas dos soluciones populares al problema de la decodificación, la de Viterbi y la secuencial, se muestra en la figura 48-124. En las curvas se comparan la decodificación de Viterbi (decisiones inflexible y flexible con valores nominales de $1/2$ y $1/3$) y la decodificación secuencial (valores nominales de $1/2$ y $1/3$). En la figura 48-124 se ilustra que para $P_B = 10^{-5}$ es posible lograr ganancias de codificación de 7 dB aproximadamente con los decodificadores secuenciales.

Desde que Shannon predijo el potencial de 11.2 dB de ganancia de codificación, parece que ya se ha logrado la mayor parte de lo que teóricamente se consideraba posible⁵.

Intercalación. Los códigos para los canales de satélites se diseñan de modo que combatan errores independientes; se denominan códigos para la corrección de errores aleatorios. También existen canales (p. ej., líneas telefónicas, almacenamiento en cinta magnética, enlaces por dispersión troposférica, y algunas veces canales de satélites) en los cuales las perturbaciones introducen errores que se agrupan en ráfagas. El empleo de un intercalador es una forma de mejorar las capacidades de corrección de errores aleatorios de un código, de modo que también es útil en un ambiente de ruido en ráfagas. El intercalador revuelve o "baraja" los bits codificados en un tramo de varios bloques de longitud (para códigos por bloques) y varias longitudes de restricción (para códigos convolucionales). La longitud requerida del tramo se determina con base en la necesidad de protección de errores durante alguna ráfaga específica. Los detalles del patrón de redistribución de bits deben ser conocidos por el receptor, así como por el transmisor, a fin de que la corriente de bits sea desintercalada antes de ser decodificada. El resultado global es una "distribución" del efecto del ruido en ráfagas, de modo que los errores inducidos parecen ser independientes (y con ello se adaptan a las capacidades de corrección de errores del código).

Ramsey³⁰ analiza configuraciones de intercaladores que reordenan una sucesión de símbolos de modo que ninguna sucesión continua de n_2 símbolos en la sucesión reordenada contiene ningún par de símbolos que estuvieron separados por menos de n_1 símbolos en el orden original. También demuestra que tal configuración es óptima en el sentido de una demora de codificación mínima posible y de la capacidad de almacenamiento.

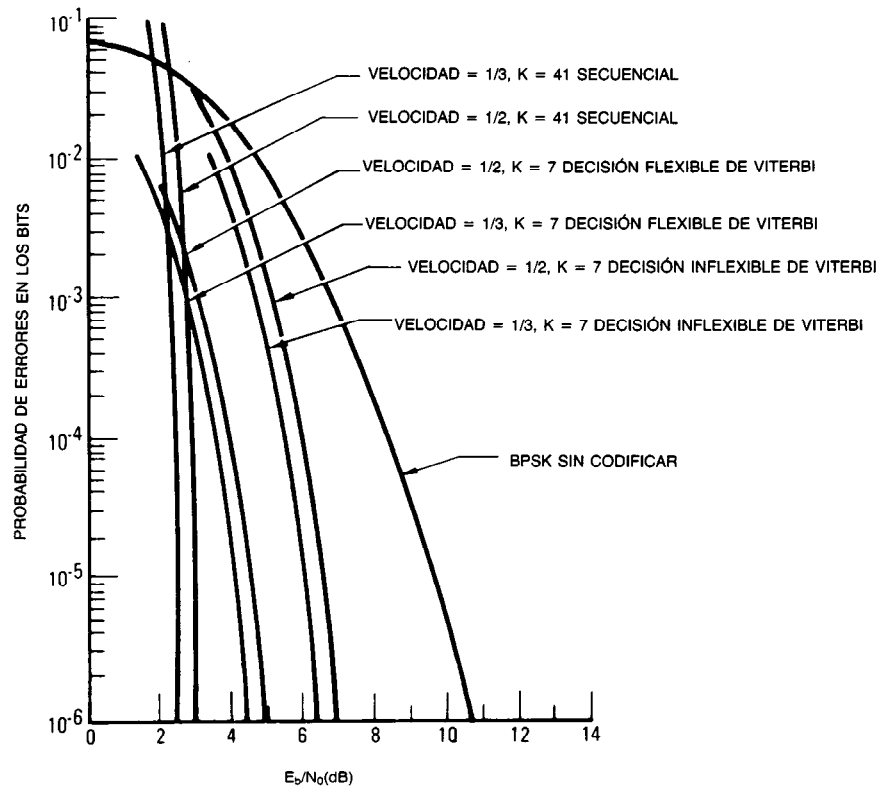


Fig. 48-124. Rendimiento de los decodificadores de Viterbi y secuencial. Manipulación por defasamiento binaria (BPSK).

Multicanalización y acceso múltiple

Los términos "multicanalización" y "acceso múltiple" se refieren al hecho de compartir un recurso de comunicación (CR, de *communication resource*) (figs. 48-111 y 48-112). Existe una diferencia sutil entre ellos. Con la multicanalización, el controlador del sistema (que puede ser una persona, un algoritmo, o incluso un tablero lógico alambrado, ya sea centralizado o distribuido) tiene conocimiento instantáneo de todos los requisitos o planes del usuario para compartir un CR. A fin de organizar la asignación de recursos no se requieren gastos generales, y suele considerarse que el proceso se verifica dentro de los confines de un emplazamiento (p. ej., un ensamblaje o un tablero de circuitos). El acceso múltiple suele implicar el acceso remoto de un recurso; además, puede haber una cantidad de tiempo diferente de cero, necesaria para que el controlador se entere de las necesidades de cada usuario del CR. Tal tiempo constituye un impacto de gastos generales para la utilización del sistema.

Fundamentalmente, existen dos métodos para mejorar la capacidad de un satélite de tráfico de comunicaciones. Uno de ellos es buscar mejoras tecnológicas hacia una mayor EIRP (potencia radiada isotrópicamente efectiva), o para obtener más ancho de banda (existe gran interés en desarrollar la banda de 30/20 GHz para la comunicación vía satélite). El segundo

método es hacer más efectiva la asignación del CR. Este segundo método es el dominio del acceso múltiple en las comunicaciones.

El problema es asignar de forma eficiente partes del CR fijo del satélite a un gran número de usuarios que pretenden comunicarse información digital en una variedad de velocidades de transferencia de bits y de mensajes, y con diferentes requisitos de tráfico. Es necesario utilizar un mecanismo mediante el cual las múltiples señales puedan acceder al CR sin crear mutua interferencia en el proceso de detección. Para evitar esta interferencia se requiere que las señales de un canal del CR no incrementen la probabilidad de error en otro canal. Debe resultar evidente que basta la ortogonalidad de las señales en canales separados para evitar la interferencia entre usuarios. Dos señales son ortogonales si es posible describirlas en el dominio del tiempo mediante la ecuación 48-139. De forma semejante, son ortogonales si es posible describirlas en el dominio de la frecuencia por medio de:

$$\int_{-\infty}^{\infty} S_i(f) S_j(f) df = 1 \quad \text{para } i = j$$

$$= 0 \quad \text{en caso contrario} \quad (48-144)$$

donde las funciones $S_i(f)$ son las transformadas de Fourier de algunas ondas de forma $s_i(t)$. La canaliza-

$$\text{Segmento } (n,m) = [t : n\tau_f + (m-1)\tau_f/M \leq t \leq n\tau_f + m\tau_f/M]$$

Otros esquemas de acceso útiles para la comunicación vía satélite son el acceso múltiple por división de espacio (SDMA, de *space division multiple access*) y el acceso múltiple por división de polarización (PDMA, de *polarization division multiple access*). Para producir SDMA, las señales de diferentes canales (a

los que se permite ocupar la misma banda de frecuencias) se transmiten mediante antenas de haces puntuales. Los haces puntuales producen ortogonalidad al separar físicamente las señales, de modo que sea posible recogerlas con receptores separados físicamente. Para la producción de PDMA, las antenas se polarizan ortogonalmente a fin de separar los campos electromagnéticos. En una implementación flexible de la SDMA, denominada TDMA con conmutación en el satélite (SS/TDMA, de *satellite-switched TDMA*), se utiliza una matriz de conmutación de microondas en el satélite. La secuencia de conmutación de la matriz se controla con una memoria programable; las señales TDMA se conectan cíclicamente con haces puntuales de diferentes antenas en secuencia rápida. Una estación terrestre en la red se comunica con las señales de otros haces transmitiendo ráfagas TDMA en sincronización adecuada con la secuencia³².

Los esquemas de acceso múltiple analizados hasta el momento podrían denominarse de asignación fija en el caso en que un usuario tiene acceso al canal independientemente del tráfico actual de mensajes. En contraste, los esquemas de asignación dinámica, algunas veces denominados acceso múltiple de asignación según demanda (DAMA, de *demand assignment multiple access*), proporcionan al usuario acceso al canal sólo cuando se tiene un mensaje que enviar. Si el tráfico de los usuarios tiende a ser del tipo en ráfagas o intermitente, es posible lograr grandes eficiencias mediante el empleo de procedimientos DAMA para el acceso al CR. En el satélite INTELSAT IV se implementó un esquema DAMA, denominado equipo de acceso según demanda, de un solo canal por portadora (SPADE, de *single channel per carrier access-on-demand equipment*) a principios de la década de 1970. En cada terminal el subsistema SPADE responde a las peticiones de servicio asignando al usuario una frecuencia portadora sin utilizar; luego avisa a las demás terminales sobre el empleo de la frecuencia por medio de un canal de señalización común. La terminal de inicio solicita un par de frecuencias al azar, lo que hace improbable que dos terminales a la vez pidan el mismo

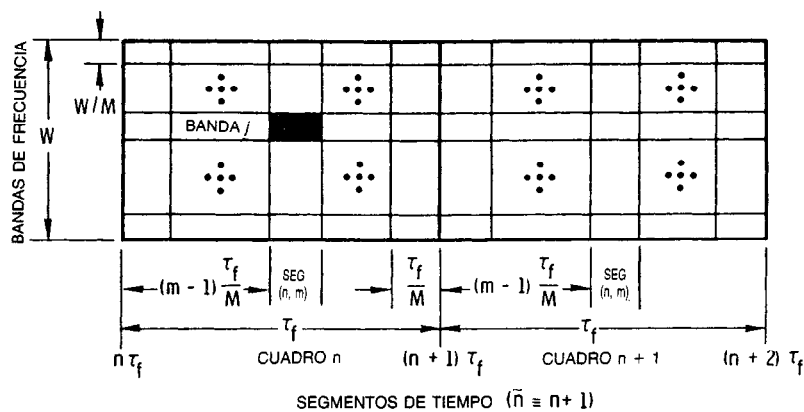


Fig. 48-125. Recurso de comunicación: diagrama de la frecuencia con respecto al tiempo de la canalización. SEG. = segmento.

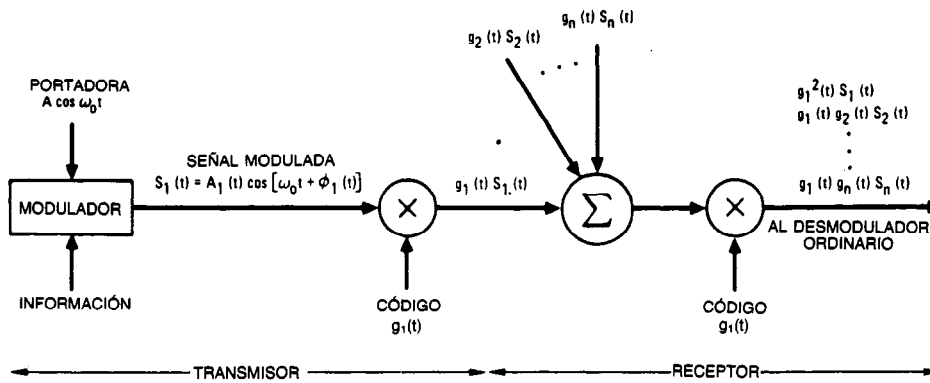


Fig. 48-126. Sistema de espectro extendido.

canal, a menos que queden muy pocos canales. En SPADE se utiliza modulación QPSK a 32 ks/s (kilosímbolos por segundo), y un ancho de banda de la señal de 38 kHz por cada canal digitalizado del habla de 64 kb/s ($R/W = 1.68$ b/s/Hz). Este primer uso comercial importante vía satélite de un esquema DAMA ha dado por resultado una utilización más eficiente de la potencia y el ancho de banda por canal que cualquier uno de los esquemas fijos de acceso múltiple³³.

El desarrollo de técnicas de conmutación por paquetes representa un avance importante para compartir los recursos de comunicaciones. En redes de conmutación por circuitos, tales como las telefónicas, las peticiones y ruta del mensaje se establecen antes de comenzar su transmisión. Una vez establecida la ruta, el mensaje se transmite por el circuito adecuado; cuando ha terminado la llamada se desconecta el circuito. En las comunicaciones por paquetes, los mensajes se agrupan en paquetes (se dividen en grupos modulares, cada uno de los cuales contiene un encabezado de la dirección). Puede considerarse como si cada paquete se moviera de forma independiente a través de la red, haciendo cola en puntos nodales específicos, junto con otros paquetes de otros tráficos. La característica clave del sistema de conmutación por paquetes es el potencial de un empleo bastante eficaz de una red de comunicaciones o de computación, especialmente en presencia de tráfico en ráfagas (de altos picos sobre el promedio). Bhargava y colaboradores⁵ presentan un resumen conciso de las características de funcionamiento y diferentes métodos de acceso que caracterizan a las redes por paquetes vía satélite.

Para las comunicaciones por medio de satélites, una consideración importante del diseño es la selección de las técnicas de señalización adecuadas para el acceso múltiple a un satélite repetidor de banda ancha con limitación fija; existen muchas opciones de esquemas de acceso múltiple³⁴. Las referencias 35 y 36 son buenas orientaciones sobre el tema del acceso múltiple para sistemas vía satélite. Nirenberg y Rubin³⁵ presentan una relación especialmente interesante entre la demora del mensaje y la densidad de potencia sobre ruido de la portadora, como función de la proporción de errores en los bits.

Dispersión de frecuencias

Las técnicas de dispersión del espectro o de espectro extendido (figs. 48-111 y 48-112) permiten que varias señales ocupen el mismo ancho de banda de RF por transmitir simultáneamente sin que haya interferencias entre ellas. La técnica se utiliza para aplicaciones tales como dar privacidad, encubrimiento de señales, rechazo de interferencia, mediciones de demora o de alcance, direccionamiento selectivo y acceso múltiple (CDMA)³⁷. En la figura 48-126 se ilustra un sistema de espectro extendido en su forma más general. Se muestra una portadora dada por $A \cos \omega_0 t$ modulada con información para producir una señal $s_1(t)$, donde:

$$s_1(t) = A_1(t) \cos[\omega_0 t + \phi_1(t)]$$

No se ha impuesto restricción alguna sobre el tipo de modulación que puede utilizarse. Ahora $s_1(t)$ está multiplicada por alguna función codificadora $g_1(t)$. (A menudo cada función codificadora se mantiene en secreto, y su empleo está restringido a una comunidad de usuarios autorizados.) La señal resultante $g_1(t)s_1(t)$ se transmite por el canal. Al mismo tiempo, otros usuarios han multiplicado sus señales por otras funciones de codificación. La señal presente en el receptor es la combinación lineal de los productos de cada usuario:

$$g_1(t)s_1(t) + g_2(t)s_2(t) + \dots + g_n(t)s_n(t) \quad (48-145)$$

La multiplicación de $s_1(t)$ por $g_1(t)$ produce una señal cuyo espectro es la convolución de los espectros de las dos señales componentes. Por tanto, si la señal $s_1(t)$ es de banda relativamente estrecha comparada con la señal de codificación o dispersión $g_1(t)$, entonces el producto tendrá casi el ancho de banda de $g_1(t)$. Supóngase que la configuración del receptor es tal que puede recibir mensajes del usuario número 1. La primera etapa del receptor multiplica la señal entrante de la ecuación 48-145 por $g_1(t)$. La salida del multiplicador producirá los siguientes términos:

Señal deseada

$$g_1^2(t)s_1(t)$$

$$g_1(t)g_2(t)s_2(t) + g_1(t)g_3(t)s_3(t) + \dots + g_1(t)g_n(t)s_n(t) \quad (48-146)$$

Si las funciones de codificación $g_i(t)$, donde $i = 1, 2, \dots, n$, se eligen con propiedades ortogonales semejantes a la ecuación 48-139, entonces la señal deseada puede extraerse perfectamente, y las señales indeseadas que producen términos cero se rechazan con facilidad.

En la figura 48-127a se ilustra la entrada de banda ancha al receptor; consta de las señales deseadas e indeseadas, cada una dispersa por su propio código, con velocidad de codificación R_p , y cada una con un espectro de forma $(\sin^2 x)/x^2$. En la figura 48-127b se ilustra el espectro después de la correlación con el código $g_1(t)$ (desdispersión). Las señales indeseadas de la ecuación 48-146 permanecen dispersas efectivamente por $g_1(t)g_i(t)$, donde $i = 2, 3, \dots, n$. Solamente la porción del espectro de las señales indeseadas que cae en el ancho de banda de la información del receptor producirá interferencia con la señal deseada.

Si existe cualquier señal perturbadora en el receptor (intencional o no), la señal de dispersión la afectará exactamente como lo hizo con la señal original en el transmisor. Por tanto, incluso una señal perturbadora de banda estrecha en medio de la banda de información será dispersada al ancho de banda de la señal de dispersión; denominémosla W . Si la potencia de la señal de dispersión es J watts, entonces su densidad promedio puede tratarse como ruido de banda ancha $J_0 = J/W$ watts por hertz. Si $s_1(t)$ tiene S watts de potencia y si la velocidad de transmisión de los datos es R bits por segundo, entonces la energía recibida por bit es $E_b = S/R$ watt-segundo y, de forma parecida a como ocurre en la ecuación 48-138, el parámetro E_b/J_0 que determina el rendimiento en cuanto a proporción de errores en los bits en presencia de perturbación o interferencia por ruido de banda ancha puede expresarse como:

$$E_b/J_0 = (S/J)(W/R)$$

También hay ruido térmico, pero se supondrá que es mucho menor que el ruido de perturbaciones o interferencia y por tanto es posible despreciarlo. Por consiguiente, la razón de potencia de perturbación entre potencia de la señal es:

$$J/S = (W/R)/(E_b/J_0) \quad (48-147)$$

Esto ilustra el hecho de que si E_b/J_0 es la mínima relación de energía de los bits sobre densidad de ruido de perturbaciones necesaria para soportar una proporción dada de errores en los bits, y si W/R es la razón de ancho de banda de la dispersión sobre la velocidad original de transferencia de los datos, también denominada ganancia de procesamiento, entonces J/S es la máxima razón tolerable de potencia de perturbación sobre potencia de la señal. Suele utilizarse como un factor de mérito para describir la vulnerabilidad de un sistema a las perturbaciones; cuanto más grande sea J/S , mayor será la resistencia a las perturbaciones. Otra forma de describir la relación de la ecuación 48-147 es la que se indica a continuación: Supóngase que un adversario quisiera utilizar una estrategia de perturbación intencional de modo que el E_b/J_0 efectivo sea lo más grande posible. Podría tratar de utilizar interferencia por pulsos, tono o banda parcial en vez de interferencia por ruido de banda ancha. Un E_b/J_0 grande podría producir un J/S pequeño para una ganancia de procesamiento fija, o podría forzar a que el comunicador utilizara una mayor ganancia de procesamiento para algún J/S deseado. El diseñador del sistema se esfuerza por conseguir una onda de forma tal que el emisor perturbador no pueda obtener alguna ventaja especial con una estrategia de congestiónamiento que no sea por ruido de banda ancha.

Existen dos técnicas populares para la dispersión del espectro (fig. 48-112). La primera se denomina dispersión del espectro por control secuencial directo o por seudoruido. La dispersión se logra multiplicando los

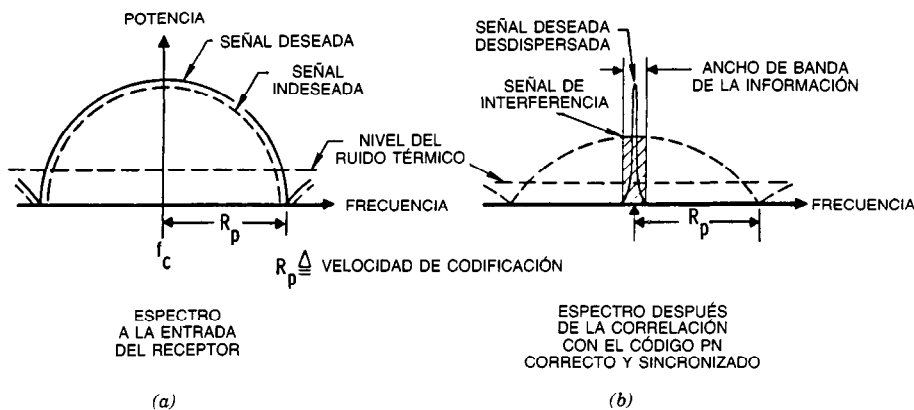


Fig. 48-127. Detección del espectro extendido: a) entrada de banda ancha al receptor, con señales indeseada y deseada; b) espectro después de la dispersión.

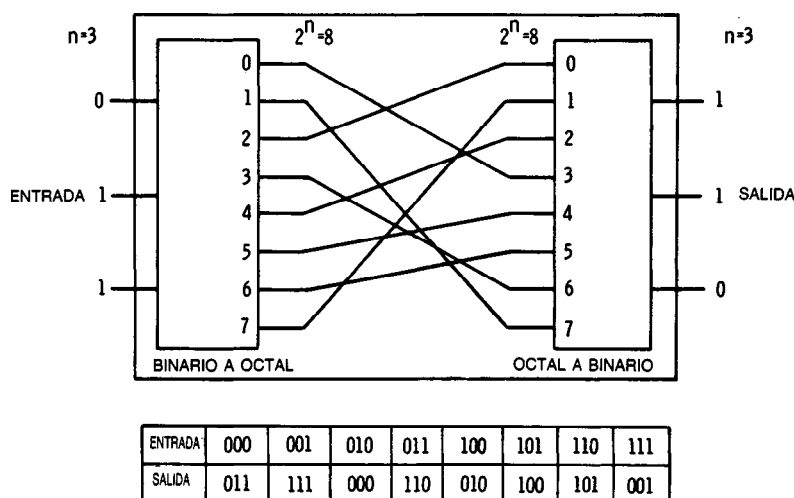


Fig. 48-128. Caja de sustitución.

datos por una sucesión binaria pseudoaleatoria (que se analiza a continuación), cuya velocidad de transferencia de los símbolos es varias veces la velocidad de transferencia de los datos. En la segunda técnica se utiliza una portadora que cambia de frecuencia. La portadora permanece a una frecuencia dada durante un instante y luego salta a la nueva frecuencia en alguna parte del ancho de banda de dispersión W . Los saltos de frecuencia suelen clasificarse como lentos o rápidos. Para el caso de los primeros, por lo general se tienen varios símbolos por salto, y el ancho de banda de la señal transmitida es igual al de la señal de los datos. En el caso de los saltos rápidos, por lo general existen varios saltos por símbolo, y el ancho de banda de la señal transmitida es igual al recíproco de la duración del salto. La figura 48-127, la ilustración espectral de un fenómeno de dispersión-desdispersión, es una interpretación exacta para cada una de las técnicas de dispersión descritas. Una diferencia importante entre las señales de control secuencial directo y las de saltos de frecuencia es que aquéllas pueden desmodularse coherentemente. Sin embargo, con los saltos de frecuencia es difícil mantener la coherencia de fase; por tanto, suelen desmodularse incoherentemente. En las referencias 38 a 41 se aborda el funcionamiento de los sistemas de dispersión del espectro.

Cifrado

Las dos razones para utilizar criptosistemas en las comunicaciones son (figs. 48-111 y 48-112):

1. La privacidad para evitar que personas no autorizadas obtengan información del canal.
2. La autenticación para evitar que personas no autorizadas introduzcan información en el canal.

El mensaje, o texto en lenguaje claro P , se cifra con una transformación reversible E_k que produce el texto cifrado $C = E_k(P)$. El texto cifrado se transmite a

través de un canal inseguro o público. Cuando un receptor autorizado recibe C , lo descifra con la transformación inversa $D_k = E_k^{-1}$ a fin de obtener:

$$D_k(C) = E_k^{-1}[E_k(P)] = P$$

que es el mensaje o texto en lenguaje claro original (sin cifrar).

E_k se elige de una familia de transformaciones criptográficas, que suelen considerarse como información pública. El parámetro k , o la clave que selecciona la transformación individual de la familia, se mantiene en secreto; en los criptosistemas típicos, cualquiera que tenga acceso a la clave es capaz de cifrar y descifrar mensajes. La clave se transmite a la comunidad de usuarios autorizados por medio de un canal adecuado (seguro, en algunos casos, un mensajero), y en general permanece sin cambio durante un número considerable de transmisiones. La meta de un escuchador oculto o adversario (criptoanalista) es producir una estimación del texto en lenguaje claro \hat{P} mediante el análisis del texto cifrado obtenido del canal público, sin contar con la clave.

Los esquemas de cifrado se clasifican en dos grandes categorías genéricas: el cifrado por bloques y el cifrado por corrientes de datos. En el cifrado por bloques, el texto en lenguaje claro se divide en varios bloques de tamaño fijo; cada bloque se cifra independientemente de los demás. Así, un bloque específico de texto en lenguaje claro se cifrará en el mismo bloque cada vez que aparezca (como en la codificación por bloques). En general, sin embargo, las propiedades deseadas en un cifrado por bloques son bastante diferentes de las deseables en un código de corrección de errores. Por ejemplo, con el cifrado, los datos del texto en lenguaje claro jamás deben aparecer en el texto cifrado; asimismo, el cambio de incluso un solo bit ya sea en el texto en lenguaje claro o en la clave provocará un cambio en aproximadamente el 50 % de los bits del texto cifrado.

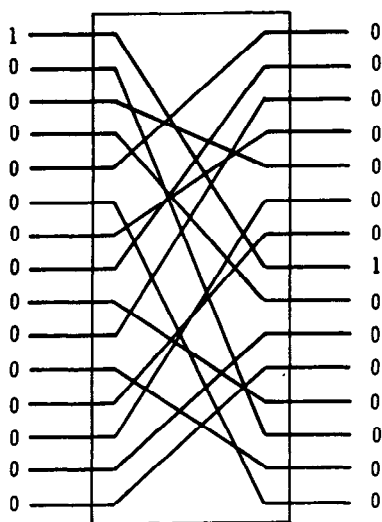


Fig. 48-129. Caja de permutación.

Los criptosistemas tienen sus raíces en el trabajo de Shannon⁴², que relaciona la criptografía con la teoría de la información. Shannon introdujo los términos "confusión" (sustitución) y "difusión" (permutación). Sugirió un método para utilizar ambos términos (un producto cifrado) a fin de construir un sistema de cifrado más robusto del que podría producir cualquiera de los métodos por sí solo. En la figura 48-128 se muestra un ejemplo de transformación no lineal por sustitución. En general, n bits de entrada se representan primero como uno de 2^n caracteres diferentes (en este ejemplo traducción de binario a octal). Luego se sustituye a uno de los demás caracteres del conjunto de 2^n caracteres, y después el carácter se convierte de regreso a una salida de n bits. Es fácil demostrar que existen $(2^n)!$ sustituciones diferentes o patrones de conexión posibles. La tarea del criptoanalista se vuelve computacionalmente imposible a medida que n se hace grande; por ejemplo, si $n = 128$, entonces $2^n = 10^{38}$, y $(2^n)!$ es un número astronómico. Se reconoce que para $n = 128$, esta caja de sustitución (caja S) representa el dispositivo ideal para cifrado. Sin embargo, aunque es posible identificar la caja S con $n = 128$ como ideal, su implementación no es factible debido a que requiere una unidad con $2^n = 10^{38}$ conexiones.

La figura 48-129 representa un ejemplo de permutación de datos (una operación lineal). Aquí los datos de entrada simplemente se reordenan o permutan (caja P). Cuando se utiliza sola, esta técnica presenta una desventaja primordial: es vulnerable a mensajes capciosos. En la figura 48-129 se ilustra uno de tales mensajes. Un solo 1 en la entrada y todo lo demás dígitos iguales a 0 revela con rapidez una de las conexiones internas; es posible utilizar mensajes similares para deducir cada una de las conexiones restantes. El producto cifrado de la figura 48-130 ilustra la combinación de las transformaciones por sustitución o permutación sugeridas originalmente por Shannon.

Representa la solución intermedia a las dificultades con la caja S o con la caja P por sí solas; la combinación de las cajas S y P produce un sistema más poderoso que el que se obtendría con cualquiera de ellas solamente. El concepto básico de producto cifrado fue utilizado por la IBM en el desarrollo de su sistema Lucifer. En 1977, la National Bureau of Standards adoptó un sistema Lucifer modificado como norma nacional estadounidense para el cifrado de datos⁴³.

En los cifrados de corrientes de datos no se tratan de forma independiente los símbolos de llegada; el cifrado depende del estado interno del dispositivo de implementación (registro de desplazamiento con retroalimentación). Después de que se ha cifrado cada símbolo, el dispositivo cambia de estado conforme a alguna regla. Por consiguiente, en general dos ocurrencias de la misma entrada de texto en lenguaje claro no se cifrarán en el mismo cifrado. En las técnicas de cifrado por corrientes de datos, por lo general, se utilizan registros de desplazamiento para generar sus secuencias pseudoaleatorias de claves. Tales secuencias se llaman así por el hecho de que aparecen aleatorias al observador casual; presentan propiedades estadísticas parecidas al experimento de lanzar al azar una moneda al aire. Sin embargo, por supuesto que no son aleatorias; están estructuradas de forma precisa, como deben estarlo a fin de poder utilizarlas para la codificación y decodificación. Un registro de desplazamiento puede convertirse en un generador de secuencias pseudoaleatorias mediante la inclusión de un ciclo de retroalimentación que calcule un nuevo término para la primera etapa con base en los n términos previos. En la figura 48-131 se muestra un ejemplo, donde $n = 4$, y la retroalimentación de las etapas 3 y 4 se suma el módulo 2 y se regresa a la etapa 1. Si la etapa inicial del registro es 1 0 0 0, entonces la sucesión de etapas activada mediante pulsos de reloj sería 1 0 0 0, 0 1 0 0, 0 0 1 0, 1 0 0 1, 1 1 0 0, ... La secuencia de salida está constituida por los bits desplazados a partir del cuarto registro; es decir, 0 0 0 1 0 0 1 1 0 1 0 1 1 1 1. Dado cualquier registro lineal de desplazamiento con retroalimentación de grado n , la secuencia de salida termina siendo periódica. Cualquier secuencia de salida que obtenga el máximo periodo posible, $p = 2^n - 1$, se denomina secuencia de longitud máxima del registro de desplazamiento²⁵.

Estas secuencias tienen las siguientes propiedades de aleatoriedad:

1. **Propiedad de equilibrio.** En cada periodo de la secuencia, el número de unos difiere del número de ceros cuando mucho en 1.
2. **Propiedad de serie.** Entre las series de unos y ceros en cada periodo, la mitad de las series de cada clase tienen longitud 1, la cuarta parte tienen longitud 2, la octava parte tienen longitud 3, etcétera.
3. **Propiedad de correlación.** Si un periodo de la secuencia se compara término a término con cualquier desplazamiento cíclico de ella misma, el número de coincidencias difiere del número de no coincidencias, a lo sumo, en 1.

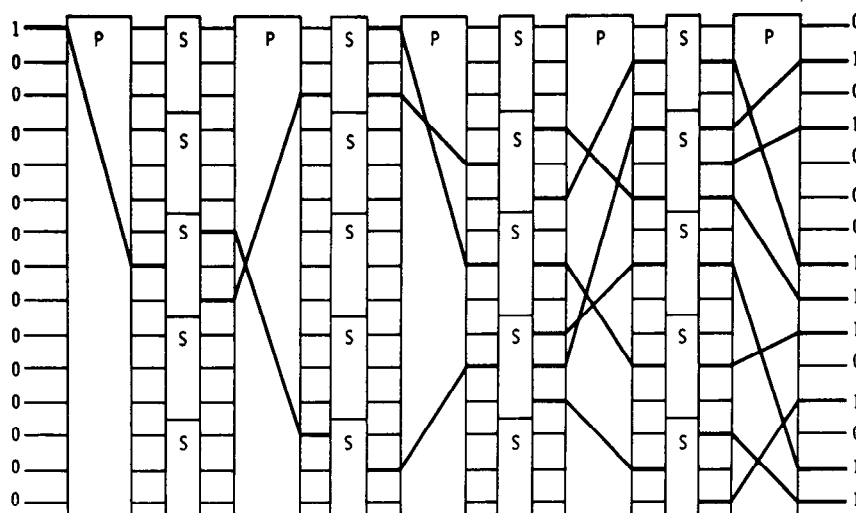


Fig. 48-130. Sistema de producto cifrado.

Los códigos lineales de longitud máxima se utilizan no sólo para las claves compactas de cifrado-descifrado, sino también para muchos propósitos parecidos, como secuencias de códigos para esquemas CDMA y otras técnicas de dispersión de frecuencia. Un registro lineal de desplazamiento es muy vulnerable al ataque de un criptoanalista. Incluso si se desconocen las derivaciones de retroalimentación, el analista solamente requiere $2n$ bits de texto en lenguaje claro y de texto cifrado para conocer tales derivaciones, el estado inicial del registro y toda la secuencia del código⁴⁴. El empleo de retroalimentación no lineal en el registro de desplazamiento dificulta la tarea del criptoanalista, si no imposible computacionalmente.

En su artículo, Diffie y Hellman⁴⁴ hacen un tratamiento exhaustivo de las bases del cifrado y descifrado; su bibliografía y referencias constituyen un recurso excelente para profundizar en el estudio del tema. También Kahn⁴⁵, presenta una historia extensa y fascinante de la criptografía desde la antigüedad hasta nuestros días. El número especial del *IEEE Communications Magazine* dedicado a la privacidad de las comunicaciones constituye otra fuente de consulta inestimable⁴⁶.

Sincronización

La sincronización puede definirse como el alineamiento de las escalas de tiempo de procesos periódicos separados espacialmente. En el contexto de las comuni-

caciones digitales, implica la estimación del tiempo y la frecuencia. Aquí solamente se dispone de espacio para enumerar y describir en términos generales los requisitos de sincronización para los sistemas digitales. A lo largo de todos los análisis de procesamiento de señales en este capítulo, se ha tratado con la operación sobre un símbolo digital m_i o una onda de forma digital $s_i(t)$ durante el intervalo de tiempo $[nT, (n+1)T]$, donde $n = 0, 1, 2, \dots$ es un entero que sirve como índice para la duración de cada símbolo, T , e $i = 1, \dots, M$ es un entero que sirve como índice para los símbolos u ondas de forma individuales de un conjunto finito. Una hipótesis implícita para cada una de las etapas del procesamiento (fig. 48-111) ha sido que el sistema está sincronizado con respecto al tiempo y la frecuencia; que el desmodulador "sabe" cuándo empezar a acumular energía para el proceso de toma de decisiones; cuándo dejar de acumularla, cuándo tomar su decisión; y cuándo repetir la operación. Cualquier error en la coordinación temporal (temporización) o en la frecuencia da por resultado una pérdida de energía, lo cual reduce efectivamente la E_b/N_0 recibida y por consiguiente degrada el rendimiento en P_B .

La jerarquía de los niveles de sincronización del sistema (fig. 48-112) es como sigue: la sincronización de la portadora se refiere a su restitución (exacta con respecto a frecuencia y fase) a partir de una onda de forma con supresión de la portadora. Solamente es necesaria para la desmodulación de sistemas coherentes

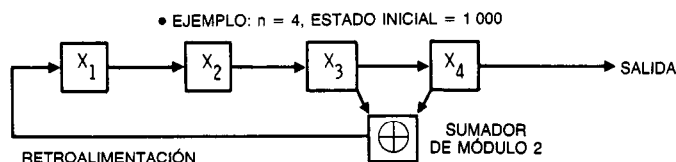


Fig. 48-131. Generador de sucesiones (secuencias) mediante registros de desplazamiento.

en fase. Se requiere la sincronización de símbolos (o bits) para determinar cuándo puede cambiar la modulación de estado. La sincronización de palabras, cuadros o paquetes se necesita para la reconstrucción adecuada de los datos. La sincronización de la red se requiere para sincronizar el acceso al canal entre varios usuarios que comparten el CR.

El diseño eficaz de la señal dicta que sea suprimida cualquier componente discreta de una portadora o señal del reloj; la potencia transmitida se dedica exclusivamente a los datos. En este caso el sistema debe recuperar la portadora y el reloj de una señal que no contenga ninguno de ellos en forma explícita. En un estudio de Franks⁴⁷ se analizan algunos sincronizadores populares y las técnicas para implementar la sincronización de símbolos y portadora. Ese estudio demuestra la aplicación de la estimación de máxima probabilidad a señales prácticas; también analiza circuitos utilizados comúnmente para aproximar estimadores de máxima probabilidad. En las referencias 3 y 48 es posible encontrar otros tratamientos detallados para implementar la sincronización de símbolos y portadora.

En la figura 48-132 se ilustran las consideraciones relativas al satélite y la terminal terrena para mantener el reloj de una terminal esclavizado al reloj del satélite (sincronización de la red). La i -ésima terminal debe ajustar sus pulsos de reloj y transmisiones de modo que lleguen sincrónicamente con los pulsos del reloj del satélite. De este modo, cualquier transmisión a partir de la i -ésima terminal debe empezar antes (con respecto al reloj del satélite) por un factor de $d_i(t)/c$, donde $d_i(t)$ es la distancia de propagación del satélite a la i -ésima terminal y c es la velocidad de la luz.

Dado que la distancia suele variar con el tiempo (ya sea que el satélite o la terminal estén en movimiento uno respecto al otro), $\tau_i(t)$, la sucesión de pulsos del reloj en la i -ésima terminal no puede mantenerse a una frecuencia de pulsos fija; por consiguiente, la duración del cuadro $\tau_i(t)$ varía con el tiempo. En muchos casos, el reloj de la terminal necesita estar esclavizado de forma continua a fin de asegurar que esté sincronizado con el reloj del satélite. La referencia 49 es una buena fuente sobre el tiempo de sincronización entre un satélite y una estación terrestre.

48.5.3 Análisis de enlaces de comunicaciones

El presupuesto de enlace de un sistema de comunicaciones es una hoja de balance de ganancias y pérdidas. Está compuesto del prorrateo detallado de los medios de transmisión y recepción, de las fuentes de ruido y de los atenuadores de señales medidos a partir del modulador y el transmisor, a través del canal, y hasta incluso el receptor y el desmodulador (fig. 48-111). El presupuesto se deduce esencialmente del cálculo de la potencia útil recibida.

Algunos de los parámetros del presupuesto son estadísticos; por ejemplo, la propagación RF se desvanece debido a eventos meteorológicos^{50,51}. Por consiguiente, el análisis del presupuesto de enlaces es una técnica de estimación para evaluar el rendimiento de los sistemas de comunicaciones.

El medio de propagación o trayectoria electromagnética que conecta al transmisor con el receptor se denomina canal. El concepto de espacio libre supone una región de canal libre de todos los objetos que pudieran afectar la propagación RF por absorción, reflexión o refracción. Además, supone que la atmósfera en el canal es perfectamente uniforme y no absorbente, y que la Tierra está infinitamente alejada o que su coeficiente de reflexión es despreciable. La energía de RF que llega al receptor se supone una función de la distancia desde el transmisor (sencillamente obedeciendo la ley de la inversa de los cuadrados de la óptica). En la práctica, por supuesto, la propagación en la atmósfera y cerca de la superficie terrestre da por resultado cierta refracción, reflexión y absorción que modifican la transmisión en el espacio libre⁵².

La SNR es una medida conveniente del rendimiento en varios puntos del enlace. La definición es:

$$SNR = \frac{\text{potencia en la onda de forma deseada}}{\text{potencia en la onda de forma de interferencia}}$$

La onda de forma deseada puede ser una señal de información, una onda de forma de la banda base, o una portadora modulada.

La SNR puede degradarse en una de dos formas posibles:

1. Por atenuación de la potencia de la onda de forma deseada con respecto a la potencia de la onda de forma de interferencia.
2. Por incremento de la potencia de la onda de forma de interferencia con respecto a la potencia de la onda de forma deseada.

Estas degradaciones se denominan "pérdida" y "ruido", respectivamente. Las pérdidas ocurren cuando, debido a algún mecanismo, una parte de la señal se desvía, dispersa o refleja de su ruta pretendida. El ruido ocurre cuando se inyecta energía de una señal indeseada al enlace, o cuando se genera ruido térmico dentro del enlace. Existen dos tipos principales de ruido: térmico y de intermodulación. El primero es radiado hacia el interior de la antena por las moléculas del oxígeno y vapor de agua presentes en la atmósfera; también es introducido por las primeras etapas del receptor. El ruido de intermodulación es producido por no linealidades en el sistema. Sus efectos dañinos suelen agruparse cuantitativamente, con otras pérdidas del sistema, bajo el encabezado de parámetro de pérdida. En las siguientes explicaciones se limitará el análisis del ruido térmico, en cuyo caso se supone que la densidad espectral de la potencia es plana en todo el intervalo de los gigahertz; el proceso del ruido térmico en los receptores de comunicaciones suele considerarse un proceso AWGN⁵³.

Un conocido modelo físico⁵⁴ para el ruido térmico o de Johnson, generado en componentes disipativos, consta de un generador de ruido con voltaje cuadrático medio de circuito abierto igual a:

$$4kT^{\circ}WR$$

donde k = constante de Boltzmann. 1.38×10^{-23} J/K

T° = temperatura, K

W = ancho de banda, Hz

R = resistencia, Ω

Puede demostrarse que la máxima potencia de ruido térmico disponible N , acoplada del generador de ruido a la sección de entrada de un amplificador, es⁵³:

$$N = kT^\circ W$$

y la densidad de ruido N_0 es sencillamente:

$$N_0 = N/W = kT^\circ \text{ W/Hz} \quad (48-148)$$

En el desarrollo de las relaciones fundamentales de la potencia del enlace se supone una fuente RF omnidireccional que transmite uniformemente sobre 4π esterradianes (radiador isotrópico). La densidad de potencia en una esfera hipotética a una distancia d de la fuente está relacionada con la potencia transmitida P_t por:

$$p(d) = P_t/4\pi d^2$$

La potencia extraída con la antenna receptora puede expresarse como:

$$P_r = p(d)A_{er} = P_t A_{er}/4\pi d^2 \quad (48-149)$$

donde el parámetro A_{er} es la sección transversal de absorción (área efectiva) de la antenna, definida por:

$$A_{er} = \frac{\text{potencia total absorbida}}{\text{densidad de flujo de potencia incidente}}$$

El área efectiva y el área física de la antenna receptora están relacionadas mediante el parámetro de eficiencia η , como sigue:

$$A_{er} = \eta A_{pr}$$

que explica el hecho de que no se absorba la potencia total; algo de ella es la pérdida por rerradiación, dispersión o desbordamiento. Dos valores típicos de η son 0.55 para un reflector parabólico y 0.75 para una bocina.

Un parámetro común de la antenna que relaciona la potencia de salida (o de entrada) con la de un radiador isotrópico es la ganancia de antenna G , donde:

$$G = \frac{\text{máxima intensidad de la potencia en alguna dirección fija}}{\text{intensidad media de la potencia en } 4\pi \text{ esterradianes}}$$

La ganancia de antenna, a diferencia de la de un amplificador electrónico, es el resultado de concentrar el flujo de RF en alguna región restringida menor que 4π esterradianes; por consiguiente, la potencia radiada efectiva con respecto a una fuente isotrópica (EIRP) se define como:

$$\text{EIRP} = P_t G_t$$

donde G_t es la ganancia de la antenna de transmisión. Para determinar la potencia recibida en el caso general en el que la fuente de transmisión manifiesta ganancia de la antenna sobre la isotrópica, se sustituye P_t por EIRP en la ecuación 48-149:

$$P_r = \text{EIRP} A_{er}/4\pi d^2 \quad (48-150)$$

La relación entre ganancia de la antenna y área efectiva de antenna es⁵⁵:

$$G = 4\pi A_e/\lambda^2 \text{ para } A_e \gg \lambda^2 \quad (48-151)$$

donde λ es la longitud de onda de la radiación. Expresiones semejantes se aplican a antenas de transmisión y recepción por medio del teorema de reciprocidad⁵⁵. Ya que el efecto de una antenna puede expresarse como una ganancia o un área, es posible sustituir A_{er} de la ecuación 48-150 por $G_r \lambda^2/4\pi$ de la ecuación 48-151, como sigue:

$$P_r = \text{EIRP} G_r \lambda^2/(4\pi d)^2 = \text{EIRP} G_r/L_s \quad (48-152)$$

En la ecuación 48-152 se han sustituido los parámetros $(4\pi d/\lambda)^2$ por el término L_s , que es la pérdida de propagación o de trayectoria. Esta pérdida caracteriza el decremento en la potencia recibida como función de la distancia y de la frecuencia; es una definición basada en el uso de una antenna receptora isotrópica ($G_r = 1$). Por tanto, la pérdida de trayectoria es una abstracción que no es posible medir; representa una pérdida de potencia recibida hipotética, que *podría ocurrir si la antenna receptora fuese isotrópica*. En un sistema de radiocomunicación, la pérdida de trayectoria explica la mayor pérdida en la potencia de la señal. En sistemas por satélite, la pérdida de trayectoria para un enlace de la banda C (6 GHz) con un satélite sincrónico suele ser de 200 dB.

Al evaluar el rendimiento de un sistema, la cantidad más importante no es la potencia recibida P_r , sino la SNR. Esto se debe a que la restricción fundamental del sistema es la capacidad de detectar la señal, con una P_B

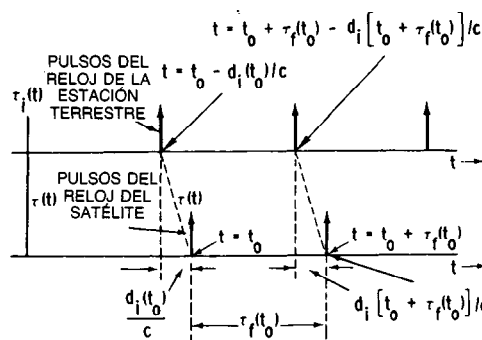


Fig. 48-132. Sincronización en el satélite y en la terminal esclava.

aceptable, en presencia de ruido. Ya que aquí la señal deseada es una onda de forma portadora modulada, a menudo se habla de una razón de potencia promedio sobre ruido de la portadora (C/N , o CNR), o (P_r/N) como la SNR de interés particular. En la ecuación 48-152 se introduce P_r/N :

$$\frac{P_r}{N} = \frac{\text{EIRP } G_r/N}{L_s} \quad (48-153)$$

La ecuación 48-153 se aplica a cualquier enlace unidireccional RF por satélite. En los sistemas analógicos, el ancho de banda del ruido suele ser mayor que el de la señal, y P_r/N es el parámetro principal para medir la detectabilidad de la señal y la calidad del funcionamiento. Sin embargo, en los receptores digitales suelen utilizarse correlacionadores o filtros de acoplamiento (compensación) en los que el ancho de banda de la señal y el del ruido se consideran iguales. En vez de considerar la potencia del ruido de entrada, un planteamiento común para los enlaces digitales es sustituir la potencia del ruido por la densidad del ruido. Puede utilizarse la ecuación 48-148 para reexpresar la ecuación 48-153 como:

$$\frac{P_r}{N_0} = \frac{\text{EIRP } G_r/T^\circ}{\kappa L_s L_o} \quad (48-154)$$

donde la temperatura efectiva del sistema T° es función del ruido térmico radiado hacia la antena y del ruido térmico generado en las primeras etapas del receptor⁵⁶⁻⁵⁸. En la ecuación 48-154 se ha introducido el término L_o para representar todos los factores de degradación debidos a diferentes pérdidas y fuentes de ruido; este término representa "otras pérdidas" no consideradas específicamente por los otros términos de la ecuación 48-153. Permite considerar un gran surtido de diversas pérdidas y fuentes de ruido (p. ej., ruido de intermodulación), que se han catalogado en detalle⁵⁹. La ecuación 48-154 resume los parámetros clave para cualquier análisis de enlaces, que son la razón de potencia de la señal recibida sobre densidad de ruido (P_r/N_0), la magnitud de la potencia transmitida (EIRP), la sensibilidad del receptor (G_r/T°) y las pérdidas ($L_s L_o$).

Si se supone que toda la potencia recibida se deriva de la señal de modulación (portadora suprimida), entonces es posible expresar, con base en la ecuación 48-138:

$$P_r/N_0 = S/N_0 = (E_b/N_0)R \quad (48-155)$$

Si algo de la potencia recibida permanece en la portadora (una pérdida de potencia de la señal), todavía es posible utilizar la ecuación 48-155, pero la potencia de la portadora se representa de forma adicional como una pérdida (dentro del parámetro L_o de la ecuación 48-154).

Hasta el momento sólo se ha hecho referencia a un tipo de E_b/N_0 , la del valor necesario de energía de los bits sobre densidad de ruido para producir una P_B específica. Ahora, a fin de facilitar el cálculo de un mar-

gen del factor de seguridad M , es necesario distinguir entre la E_b/N_0 necesaria y la E_b/N_0 verdadera o recibida. A partir de este momento, la primera se denominará $(E_b/N_0)_{\text{reqd}}$, y la segunda, $(E_b/N_0)_r$. La ecuación 48-155 puede reexpresarse, introduciendo el parámetro M del margen de enlace, como:

$$P_r/N_0 = (E_b/N_0)_r R = M(E_b/N_0)_{\text{reqd}} R \quad (48-156)$$

La diferencia en decibels entre $(E_b/N_0)_r$ y $(E_b/N_0)_{\text{reqd}}$ produce el margen de enlace. Considérese que un sistema está especificado para funcionar a una $(E_b/N_0)_{\text{reqd}} = 10 \text{ dB}$, con $P_B = 10^{-4}$. Supóngase que se requiere un margen de enlace de 4 dB (se supondrá que el P_B medido por una E_b/N_0 de 14 dB es 10^{-6}). Es posible considerar este margen de dos formas:

1. Puede afirmarse que se tiene 4 dB más de E_b/N_0 de lo realmente necesario para cumplir la P_B necesaria de 10^{-4} .
2. Puede afirmarse que se está operando a una E_b/N_0 de 14 dB y, por consiguiente, que la P_B de operación verdadera del sistema es 10^{-6} , un margen de 100 veces mejor comportamiento de probabilidad de error de lo necesario.

El parámetro $(E_b/N_0)_{\text{reqd}}$ refleja las diferencias de un sistema a otro; tales diferencias podrían deberse a diferencias en los esquemas de modulación o codificación. Una $(E_b/N_0)_{\text{reqd}}$ mayor que la esperada podría deberse a un sistema RF subóptimo, que adolezca de grandes errores de coordinación temporal o permita más ruido en el proceso de detección que un filtro ideal equivalente.

Al combinar las ecuaciones 48-154 y 48-156 y despejando el margen de enlace M , se obtiene:

$$M = \frac{\text{EIRP } G_r/T^\circ}{(E_b/N_0)_{\text{reqd}} R \kappa L_s L_o} \quad (48-157)$$

Ya que el análisis presupuestario del enlace suele calcularse en decibels, es posible expresar la ecuación 48-157 como:

$$M(\text{dB}) = \text{EIRP}(\text{dBW}) + G_r(\text{dBI}) - (E_b/N_0)_{\text{reqd}}(\text{dB}) - R(\text{dB} - \text{b/s}) - \kappa T^\circ(\text{dBW/Hz}) - L_s(\text{dB}) - L_o(\text{dB}) \quad (48-158)$$

La potencia de la señal transmitida se expresa en decibels-watts (dBW), la densidad de ruido en decibels-watts por hertz (dBW/Hz), la ganancia de la antena en decibels respecto a la ganancia isotrópica (dBI), la velocidad de los datos en decibels respecto a b/s (dB-b/s), y los demás términos están en decibels (dB). El valor de los parámetros en la ecuación 48-158 constituye el cálculo presupuestario del enlace, una herramienta útil para asignar los recursos de comunicaciones. En un esfuerzo por mantener un margen positivo podría haber soluciones de compromiso entre los parámetros; podría elegirse reducir la potencia del transmisor al ceder un margen en exceso, o bien podría

elegirse incrementar la velocidad de transmisión de datos mediante la reducción necesaria de la E_b/N_0 (mediante la selección de modulación y codificación mejoradas). Cada uno de los decibeles de la ecuación 48-158, sin importar a qué parámetro pertenezca, es tan útil como cualquiera de los demás decibeles. Sin embargo, debe observarse que a medida que los requisitos se restringen más, podría no ser posible encontrar soluciones intermedias o efectuar ajustes en algunos parámetros. Por ejemplo, aun cuando la modulación PSK binaria es mejor que la FSK binaria (en el sentido de la P_B), los requisitos para operar en un ambiente centelleante podrían determinar la no utilización de la PSK y la elección de la FSK más robusta. Además, ciertos requisitos de cobertura podrían restringir las dimensiones de la antena, de modo que no se cuente con la libertad para ajustar o elegir cualquier ganancia de la antena que se desee.

¿Cuánto margen de enlace se requiere?

Suele plantearse a menudo la pregunta de con cuánto margen de enlace debe diseñarse el sistema. La respuesta es que si todas las fuentes de ganancia, pérdida y ruido se han detallado rigurosamente, y si los parámetros de enlace con grandes varianzas (p. ej., desvanecimientos debidos a condiciones climatológicas) cumplen con los requisitos estadísticos para la disponibilidad del enlace, entonces se requiere muy poco margen. Para las comunicaciones por satélite en la banda C, donde los parámetros son bien conocidos y se comportan de manera bastante aceptable, podría diseñarse un sistema con sólo 1 dB de margen de enlace. Las estaciones de televisión solamente receptoras que operan con reflectores parabólicos de 16 pies de diámetro en la banda C suelen diseñarse con un margen de menos de 1 dB. Sin embargo, las comunicaciones telefónicas por satélite en las que se utilizan normas de disponibilidad de 99.9 % requieren un margen considerablemente superior; algunos de los sistemas INTELSAT tienen 4 o 5 dB de margen. Los diseños en los que se utiliza una frecuencia superior (p. ej., 14/12 GHz) a menudo requieren márgenes mayores, debido a que las pérdidas atmosféricas aumentan con la frecuencia y son altamente variables. Debe observarse que un subproducto de la atenuación debida a la pérdida atmosférica es mayor ruido de la antena. Cuando se tolera un margen extra para la pérdida debida a condiciones climatológicas, también debe agregarse un margen adicional a fin de compensar el incremento en la temperatura de la antena (que es función del ruido térmico radiado hacia el interior de la antena). Con amplificadores de bajo ruido, pequeños cambios climatológicos pueden dar por resultado incrementos de 40 a 50 K en la temperatura de la antena.

Repetidores de señales transmitidas por satélite

Las repeticiones de señales transmitidas por satélite retransmiten los mensajes que reciben (con una traslación en la frecuencia portadora). Una repetidora regenerativa (digital) regenera —es decir, desmodula y reconstituye— la información digital contenida en las ondas de forma recibidas; sin embargo, una repetidora

no regenerativa solamente amplifica, pero no transforma la señal a su formato de banda base. De este modo, una repetidora no regenerativa puede utilizarse con muchos formatos diferentes de modulación (de forma simultánea), pero por lo general una repetidora regenerativa se diseña para funcionar con sólo uno o muy pocos formatos de modulación. En el análisis de enlace para una repetidora de señales transmitidas por satélite se tratan el enlace ascendente y el enlace descendente como dos análisis por separado de punto a punto. A fin de estimar el rendimiento de un enlace con repetidora regenerativa es necesario determinar por separado la probabilidad de errores en los bits en los enlaces ascendente y descendente. La proporción global de errores se obtiene mediante la simple adición de las proporciones individuales^{60,61}.

En contraste, en el análisis del enlace con una repetidora no regenerativa por lo general se trata el "viaje de ida y vuelta" (transmisión de enlace ascendente hacia el satélite y retransmisión de enlace descendente hacia la terminal terrestre) como un solo análisis. A fin de estimar el rendimiento de un enlace con repetidora no regenerativa, los valores E_b/N_0 (o P_s/N_0) de los enlaces ascendente y descendente se combinan como sigue, en ausencia de ruido de intermodulación⁶⁰:

$$(E_b/N_0)_U^{-1} + (E_b/N_0)_D^{-1} = (E_b/N_0)_R^{-1}$$

donde los subíndices U , D y R representan los valores del enlace ascendente, del enlace descendente y resultantes de E_b/N_0 , respectivamente.

La mayor parte de los satélites comerciales actuales pertenecen a la clase simple no regenerativa. Sin embargo, parece evidente que los sistemas digitales por satélite futuros, que requerirán procesamiento, conmutación o direccionamiento selectivo de mensajes a bordo, empezarán con una repetidora regenerativa para transformar las ondas de forma recibidas en bits de mensajes.

Además del potencial para el procesamiento complejo de datos, una de las ventajas más importantes de las repetidoras regenerativas en comparación con las no regenerativas es el desacoplamiento de los enlaces ascendente y descendente, de modo que la potencia del ruido del enlace ascendente no sea retransmitida en el enlace descendente. Se cuenta con mejoras de funcionamiento importantes en términos de reducción de los valores E_b/N_0 necesarios en los enlaces ascendente y descendente con respecto a los valores necesarios en los diseños ordinarios de transportadores usados en la actualidad. Se han demostrado mejoras hasta de 5 dB en el enlace ascendente y de 6.8 dB en el enlace descendente (utilizando modulación QPSK coherente, con $P_B = 10^{-4}$)⁶⁰.

La potencia está seriamente limitada en la mayor parte de los sistemas de comunicación por satélite, y las ineficiencias asociadas con las etapas de amplificación lineal de potencia son intolerables. Por ello, en muchas repetidoras de señales transmitidas por satélite se utilizan amplificadores de potencia bastante no lineales; la característica más importante aquí es que se

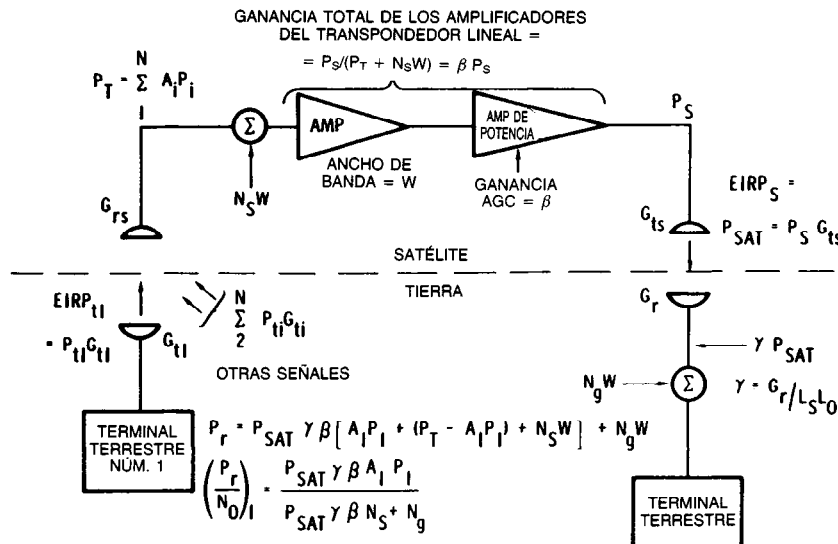


Fig. 48-133. Repetidora de satélite no regenerativa.

obtiene amplificación de potencia eficiente a expensas de distorsiones no lineales. Los efectos indeseables de las no linealidades de la repetidora son:

1. Ruido de intermodulación (IM) debido a la acción multiplicativa de diferentes portadoras. El daño provocado es doble: puede perderse potencia útil del canal como energía IM, y pueden introducirse productos espurios IM como interferencia en el canal^{62,63}.
2. La conversión de AM a PM es un fenómeno de ruido de fase que ocurre en dispositivos no lineales, tales como los tubos de ondas progresivas (viajeras) (TWT, de *traveling-wave tubes*). La fluctuación en el nivel de operación (modulación de amplitud) produce variaciones de fase que afectan el rendimiento en P_B de sistemas en los que se utiliza un formato de modulación MPSK^{64, 65}.
3. Supresión de señales débiles por señales fuertes⁶², hasta de 6 dB.

Las repetidoras no regenerativas ordinarias suelen operarse "despojadas" de su región saturada altamente no lineal, a fin de evitar el considerable ruido IM y así permitir la utilización eficiente de todo el ancho de banda del sistema. Sin embargo, el "despojo" hasta la región lineal es una solución intermedia; es necesario aceptar algo de ruido IM a fin de obtener un nivel útil de potencia de salida.

Vale la pena describir aquí un conjunto de características únicas de las repetidoras no regenerativas: la dependencia de la SNR del enlace descendente respecto a la SNR del enlace ascendente, y el compartimiento de la potencia del enlace descendente de la repetidora en proporción a la potencia del enlace ascendente de cada una de las señales y el ruido del enlace ascenden-

te. Desde ahora, la alusión a una repetidora o a un transportador implicará una repetidora no regenerativa, y por simplicidad se supondrá que el transpondedor se encuentra funcionando en su intervalo lineal.

Un transpondedor de satélite está limitado en cuanto a capacidad de transmisión por su potencia de enlace descendente, la potencia del enlace ascendente de la terminal terrestre, el ruido del satélite y la estación terrestre, y el ancho de banda del canal. Por lo común uno de estos parámetros representa la restricción dominante de funcionamiento; más a menudo se demuestra que la potencia del enlace descendente o el ancho de banda del canal son la limitación de mayor importancia del sistema. En la figura 48-133 se ilustran los parámetros importantes de enlace de un canal lineal de repetidora de señales transmitidas por satélite. La repetidora transmite todos los mensajes (o el ruido, si no hay mensajes) que llegan por el enlace ascendente, sin procesamiento alguno más allá de la traslación de la frecuencia. Supóngase que los múltiples enlaces ascendentes en el ancho de banda del receptor W se separan entre sí mediante el empleo de un esquema de acceso múltiple, tal como el FDMA o el CDMA. La potencia del enlace descendente del satélite P_{sat} es constante y, como se supone un transpondedor lineal, P_{sat} se comparte entre las múltiples señales (y el ruido) del enlace ascendente en proporción a sus niveles de potencia respectivos.

La transmisión empieza en una estación terrestre (ancho de banda $< W$), por ejemplo, en la terminal número 1, con $EIRP_{t1} = P_{t1} G_{t1}$. Simultáneamente se están transmitiendo otras señales al satélite (desde otras terminales). En el satélite se recibe una potencia de la señal total $P_T = \sum A_i P_i$, donde las A_i reflejan las diversas pérdidas de propagación que experimentan las diferentes señales al llegar al satélite. En el satélite también se recibe la potencia de ruido $N_S W$, donde N_S

es la densidad de ruido debida al ruido térmico radiado hacia la antena del satélite y el ruido térmico generado en el receptor del satélite. La $EIRP_s = P_{sat}$ total del enlace descendente del satélite puede expresarse mediante la siguiente identidad⁴:

$$P_{sat} = P_{sat}\beta [A_i P_1 + (P_T - A_i P_1) + N_s W]$$

donde $\beta = 1/(P_T + N_s W)$ es la ganancia AGC y P_T se ha expresado deliberadamente como $A_i P_1 + (P_T - A_i P_1)$ a fin de separar la potencia de la señal número 1 de las demás señales simultáneas en el transpondedor. Utilizando la ecuación 48-154, $(P_r/N_0)_{li}$ para la señal número 1 que llega a la i -ésima terminal puede expresarse como sigue⁴:

$$\left(\frac{P_r}{N_0}\right)_{li} = \frac{P_{sat}\gamma_i\beta A_i P_1}{P_{sat}\gamma_i\beta N_s + N_g} \quad (48-159)$$

donde $\gamma_i = G_r/L_s L_o$ para la i -ésima terminal y N_g es la densidad de ruido del receptor para la i -ésima terminal.

Cuando domina el ruido de la repetidora de señales transmitidas por satélite, es decir, cuando $P_T \ll N_s W$, se dice que el canal está limitado en enlace ascendente, y la mayor parte de P_{sat} del enlace descendente se asigna sin provecho a la potencia de ruido. Cuando éste es el caso y cuando $\gamma_i P_{sat} \gg N_g W$, la ecuación 48-159 puede escribirse como:

$$\left(\frac{P_r}{N_0}\right)_{li} \cong \frac{\gamma_i P_{sat} A_i P_1 / N_s W}{(\gamma_i P_{sat} / W) + N_g} \cong \frac{A_i P_1}{N_s} \quad (48-160)$$

La ecuación 48-160 ilustra que, en el caso de un canal limitado en enlace ascendente, la razón resultante P_r/N_0 obedece en lo esencial la SNR del enlace ascendente. La situación más común es el canal limitado en enlace descendente, en cuyo caso $P_T \gg N_s W$ y la $EIRP$ del satélite está limitada. En este caso, la ecuación 48-159 puede expresarse como⁴:

$$\left(\frac{P_r}{N_0}\right)_{li} \cong \frac{\gamma_i P_{sat} A_i P_1 / P_T}{N_g}$$

Así, la potencia del transpondedor es compartida esencialmente entre las diferentes señales transmitidas en enlace ascendente; al enlace descendente se transmite muy poco ruido del enlace ascendente. En este caso de limitación de enlace descendente, la P_r/N_0 resultante del repetidor está restringida solamente por los parámetros del enlace descendente.

REFERENCIAS BIBLIOGRÁFICAS

1. W. L. Pritchard, "Satellite Communication—An Overview of the Problems and Programs", *Proc. IEEE* **65**:294-307.
2. M. P. Ristenbatt, "Alternatives in Digital Communications", *Proc. IEEE* **61**:703-721.
3. W. C. Lindsey y M. K. Simon, *Telecommunication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ.
4. J. J. Spilker, Jr., *Digital Communication by Satellite*, Prentice-Hall, Englewood Cliffs, NJ.
5. V. K. Bhargava, D. Haccoun, R. Matyas y P. P. Nuspl, *Digital Communications by Satellite*, Wiley, Nueva York.
6. C. B. Feldman y W. R. Bennett, "Band Width and Transmission Performance", *BSTJ* **28**:594-595.
7. A. Lender, "The Duobinary Technique for High Speed Data Transmission", *IEEE Trans. Commun. Electron.* **82**:214-218.
8. E. R. Kretzmer, "Generalization of a Technique for Binary Data Communication", *IEEE Trans. Commun. Tech.*, pp. 67-68.
9. S. Pasupathy, "Correlative Coding: A Bandwidth Efficient Signaling Scheme", *IEEE Commun. Mag.*, pp. 4-11.
10. N. S. Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers", *Proc. IEEE* **62**(5):611-632.
11. B. S. Atal y S. L. Hanover, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *J. Acoust. Soc. Am.*, **50**:637-655.
12. J. W. Bayless, S. J. Campanella y A. J. Goldberg, "Voice Signals: Bit by Bit", *IEEE Spectr.* **10**:28-39.
13. J. L. Flanagan, *Speech Analysis, Synthesis, and Perception*, 2.^a ed., Springer-Verlag, Nueva York.
14. R. E. Crochiere y J. L. Flanagan, "Current Perspectives in Digital Speech", *IEEE Commun. Mag.* **21**(1):32-40.
15. D. Huffman, "A Method for Constructing Minimum Redundancy Codes", *Proc. IRE* **40**:1098-1101.
16. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, Nueva York.
17. S. A. Gronemeyer y A. L. McBride, "MSK and Offset QPSK Modulation", *IEEE Trans. Commun.* **COM-24**(8):809-820.
18. S. Pasupathy, "Minimum Shift Keying: A Spectrally Efficient Modulation", *IEEE Commun. Mag.*, pp. 14-22.
19. E. Arthurs y H. Dym, "On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise—A Geometric Interpretation of Three Basic Data Transmission Systems", *IRE Trans. Commun. Sys.* **CS-10**:336-372.
20. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Parte I, Wiley, Nueva York.
21. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, Nueva York.
22. K. Feher, *Digital Communications: Microwave Applications*, Prentice-Hall, Englewood Cliffs, NJ.
23. P. R. Hartman, "Digital Radio Technology: Present and Future", *IEEE Commun. Mag.* **19**(4):10-14.
24. J. M. Wozencraft e I. M. Jacobs, *Principles of Communication Engineering*, Wiley, Nueva York.

25. S. Golomb, ed., *Digital Communications with Space Applications*, Prentice-Hall, Englewood Cliffs, NJ.
26. V. K. Bhargava, "Forward Error Correction Schemes for Digital Communications", *IEEE Commun. Mag.* **21**(1):11-19.
27. A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems", *IEEE Trans. Commun. Tech.* **COM-19**:751-772.
28. J. A. Heller e I. M. Jacobs, "Viterbi Decoding for Satellite and Space Communication", *IEEE Trans. Commun. Tech.* **COM-19**:835-848.
29. G. Forney, Jr., "Coding and Its Application in Space Communications", *IEEE Spectr.*
30. J. L. Ramsey, "Realization of Optimum Interleavers", *IEEE Trans. Infor. Theory* **IT-16**(3):338-345.
31. I. L. Lebow, K. L. Jordan, Jr., y P. R. Drouilhet, Jr., "Satellite Communications to Mobile Platforms", *Proc. IEEE* **59**(2):139-159.
32. T. Muratani, "Satellite-Switched Time-Domain Multiple Access", *Record IEEE Electron. Aerosp. Syst. Conv. (EASCON)*, pp. 189-196.
33. J. G. Puente y A. M. Werth, "Demand-Assigned Service for the Intelsat Global Network", *IEEE Spectr.*, pp. 59-69.
34. J. W. Schwartz, J. M. Aein y J. Kaiser, "Modulation Techniques for Multiple Access to a Hard-Limiting Satellite Repeater", *Proc. IEEE* **54**(5):763-777.
35. L. M. Nirenberg e I. Rubin, "Multiple Access System Engineering—A Tutorial", *IEEE WESCON/78 Professional Program*, Modern Communication Techniques and Applications, Sesión 21, Los Ángeles.
36. G. D. Dill, "TDMA, The State-of-the-Art", *Record IEEE Electron. Aerosp. Syst. Conv. (EASCON)*, pp. 31-5A a 31-5I.
37. R. C. Dixon, *Spread Spectrum Analysis*, Wiley, Nueva York.
38. M. P. Ristenbatt y J. L. Daws, Jr., "Performance Criteria for Spread Spectrum Communications", *IEEE Trans. Commun.* **COM-25**:756-761.
39. S. W. Houston, "Modulation Techniques for Communication, Part I: Tone and Noise Jamming Performance of Spread Spectrum M-ary FSK and 2, 4-ary DPSK Waveforms", *IEEE 1975 Nat. Aerosp. Electron. Conf.*, pp. 51-58.
40. G. K. Huth, "Spread Spectrum Techniques", *IEEE WESCON/78 Professional Program*, Modern Communication Techniques and Applications, Sesión 15, Los Ángeles.
41. J. K. Holmes, *Coherent Spread Spectrum Systems*, Wiley, Nueva York.
42. C. E. Shannon, "Communication Theory of Secrecy Systems", *BSTJ* **28**:656-715.
43. National Bureau of Standards, "Data Encryption Standard", *Federal Information Processing Standard (FIPS)*, Publication n.º 46.
44. W. Diffie y M. E. Helman, "Privacy and Authentication: An Introduction to Cryptography", *Proc. IEEE* **67**(3):397-427.
45. D. Kahn, *The Codebreakers*, Macmillan, Nueva York.
46. Special Issue on Communications Privacy, *IEEE Commun. Mag.* **16**(6).
47. L. E. Franks, "Carrier and Bit Synchronization in Data Communication—A Tutorial Review", *IEEE Trans. Commun.* **COM-28**(8):1107-1121.
48. W. Lindsey, *Synchronization Systems in Communications and Control*, Prentice-Hall, Englewood Cliffs, NJ.
49. P. P. Nuspl, K. E. Brown, W. Seenaart y B. Ghicopoulos, "Synchronization Methods for TDMA", *Proc. IEEE* **65**:434-444.
50. R. K. Crane, "Prediction of Attenuation by Rain", *IEEE Trans. Commun.* **COM-28**(9):1717-1733.
51. L. M. Schwab, "World-Wide Link Availability for Stationary and Critically Inclined Orbits Including Rain Attenuation Effects", *Lincoln Laboratory, Project Report DCA-9*.
52. P. L. Bargellini, "Principles and Evolution of Satellite Communications", *Acta Astron.* **5**:135-149.
53. R. Gagliardi, *Introduction to Communication Engineering*, Wiley, Nueva York.
54. H. Nyquist, "Thermal Agitation of Electric Charge in Conductors", *Phys. Rev.* **32**:110-113.
55. R. E. Collin y F. J. Zucker, *Antenna Theory*, Parte I, McGraw-Hill, Nueva York.
56. P. F. Panter, *Communications Systems Design: Line-of-Sight and Tropo Scatter Systems*, McGraw-Hill, Nueva York.
57. D. C. Hogg y T-S. Chu, "The Role of Rain in Satellite Communications", *Proc. IEEE* **63**(9):1308-1331.
58. D. C. Hogg y W. W. Mumford, "The Effective Noise Temperature of the Sky", *Microw. J.*, pp. 80-84.
59. B. Sklar, "What the System Link Budget Tells the System Engineer", *Proc. Int. Telemetry Conf.* **15**.
60. S. J. Campanella, F. Assal y A. Berman, "On-board Regenerative Repeaters", *Int'l. Conf. Communications*, Chicago, vol. 1, pp. 6.2-121 a 6.2-125.
61. K. Koga, T. Muratani y A. Ogawa, "On-Board Regenerative Repeaters Applied to Digital Satellite Communications", *Proc. IEEE* **65**(3):401-410.
62. J. J. Jones, "Hard Limiting of Two Signals in Random Noise", *IEEE Trans. Infor. Theory*, pp. 34-42.
63. F. E. Bond y H. F. Meyer, "Intermodulation Effects in Limiter Amplifier Repeaters", *IEEE Trans. Commun. Tech.* **COM-18**(2):127-135.
64. O. Shimbo, "Effects of Intermodulation, AM-PM Conversion, and Additive Noise in Multicarrier TWT Systems", *Proc. IEEE* **59**:230-238.
65. P. Jain, T. C. Huang, K. T. Woo y colaboradores, "Detection of MPSK Signals Transmitted Through a Nonlinear Satellite Repeater", *NTC '77 Conference Record*.

48.6 COMUNICACIONES POR CANALES CON DESVANECIMIENTO*

Peter Monsen

48.6.1 Introducción

Actualmente vuelven a ser tema de análisis dos canales de radiopropagación para comunicaciones transhorizonte: la tropodispersión y la HF. En el pasado, la transmisión por estos canales de radio se consideraba no confiable debido a los efectos del desvanecimiento. Sin embargo, recientemente la transmisión con conversión de analógico a digital y el empleo de nuevas técnicas adaptativas de procesamiento de la señal han ofrecido una calidad promisorio aceptable de la red de comunicación. Además, la radio transhorizonte tiene ventajas económicas, de seguridad o ambas con respecto a los enlaces por satélite, cable o microondas terrestres en línea recta. Existe un interés renovado por el uso de las comunicaciones de alta frecuencia (HF) y por tropodispersión para redes portadoras de tráfico digital.

Para la radio HF se utilizan frecuencias en el intervalo de 2 a 30 MHz. A estas frecuencias, las comunicaciones transhorizonte se logran por medio de flexión refractiva de la onda de radio en la ionosfera desde capas ionizadas a diferentes altitudes. En la mayor parte de los casos más de una "capa" ionosférica provoca el regreso de una onda de radio refractada a la antena receptora. La respuesta de pulsos de tal canal presenta una estructura discreta de varias trayectorias. El tiempo entre la primera llegada y la última llegada es la dispersión de demora de las trayectorias múltiples. Los cambios en la densidad de iones en las capas individuales debidos al calentamiento solar provocan fluctuaciones en cada regreso de trayectorias múltiples. La característica de variación en el tiempo de las diversas trayectorias produce de forma alternada interferencias destructiva y constructiva. El desvanecimiento resultante es posible que produzca la pérdida completa de la señal.

La radiotransmisión por tropodispersión fue descubierta apenas terminada la Segunda Guerra Mundial, cuando se observó que las señales de microondas de radares transhorizonte eran mucho más intensas que las predichas con base en cálculos de difracción sobre la superficie de la Tierra. Una teoría común acerca de este fenómeno, desarrollada por Tatarski¹, establece que las fluctuaciones aleatorias en la constante dieléctrica en la troposfera desvían una pequeña fracción de la energía incidente de regreso al receptor. El nombre "dispersión troposférica" se deriva de este concepto de redireccionamiento aleatorio de la onda incidente por la troposfera. Ocurren importantes regresos dispersivos desde un "volumen común", definido por los patrones de los haces de las antenas receptora y transmisora. Los regresos dispersivos desde diferentes puntos dentro del volumen común tienen diferentes retardos

de trayectoria. Las señales dispersadas desde puntos separados por más de una distancia de descorrelación de las fluctuaciones en la constante dieléctrica no están correlacionadas. Así, como en un ejemplo de HF, la respuesta de pulsos que caracteriza al canal tiene una estructura de trayectoria múltiple variante en el tiempo con dispersión de retardos, pero sin las capas discretas. Los sistemas de tropodispersión se han utilizado en forma amplia en aplicaciones militares para comunicaciones transhorizonte hasta unas 600 millas. El intervalo de frecuencia aproximado para esta aplicación es de 400 a 5 000 MHz. Las comunicaciones confiables requieren trayectorias redundantes de transmisión que se obtienen mediante el empleo de frecuencias múltiples, antenas separadas en el espacio, o dispersión a dos diferentes ángulos del haz. Las diversas trayectorias redundantes se denominan trayectorias de diversidad, y el número de trayectorias se denomina orden de la diversidad.

La dispersión de retardos de trayectoria múltiple limita la capacidad del canal que es posible lograr en sistemas analógicos actuales. Solamente es posible lograr anchos de banda de transmisión menores que el recíproco de este retardo de trayectoria múltiple. Las señales de anchos de banda mayores se distorsionan debido a la dispersión de trayectoria múltiple. En los sistemas FM esta dispersión provoca ruido de intermodulación después de la detección.

Con la introducción de los sistemas de comunicación por satélite, que no adolecen de amplio desvanecimiento de trayectoria múltiple, el futuro de los sistemas HF y de tropodispersión parecía limitado. Factores económicos y de seguridad han modificado esta afirmación, especialmente para la transmisión digital. Con formatos de señal digital pueden diseñarse métodos adaptativos a fin de medir la estructura de trayectoria múltiple y explotarla como una forma extra de diversidad para mejorar el rendimiento. A diferencia de la capacidad de los sistemas analógicos, la capacidad de los sistemas digitales no está restringida por la dispersión de demoras de trayectoria múltiple. Desde el punto de vista de la red, los desvanecimientos en enlaces digitales en tándem no tienen efecto acumulativo debido a que la señal puede regenerarse en cada nodo.

Los sistemas adaptativos por tropodispersión han resultado capaces de detectar eficientemente señales distorsionadas por un medio con desvanecimiento mientras se siguen las variaciones del desvanecimiento. Si la adaptación del receptor requiere significativamente menos razón de señal sobre ruido por bit que la detección por el receptor de la señal digital, entonces las decisiones del receptor pueden utilizarse de forma eficaz como una estimación de la señal transmitida para lograr lo que se denomina adaptación *dirigida por decisiones*.

Tales sistemas pueden hacerse funcionar sin la transmisión de señales especiales piloto o de referencia para el seguimiento del canal.

En la comunicación por tropodispersión, las técnicas adaptativas han incrementado la capacidad de velocidad digital por lo menos en un orden de magnitud. Un

*Reimpreso con permiso del *IEEE Communications Magazine*, Vol. COM-28, núm. 1, págs. 16 a 25.

modem ecualizador autoadaptable^{2,3} desarrollado para enlaces troposféricos militares ha sido probado con éxito en el campo a velocidades digitales hasta de 12.6 Mb/s en una asignación de canal de 15 MHz. En los sistemas HF, las técnicas adaptativas de procesamiento de señales están siendo consideradas con objetivos de velocidades digitales del orden de 5 kb/s en un canal de 3 kHz. En estos dos ejemplos el periodo del símbolo digital es del mismo orden que la dispersión de retardos de trayectoria múltiple del canal.

La aplicabilidad de las técnicas adaptativas de procesamiento de señales depende en gran medida de si la velocidad de desvanecimiento es o no menor que la velocidad de la señalización. Como se analizará posteriormente en esta sección, ambos enlaces por radio HF y por tropodispersión pueden considerarse canales de trayectoria múltiple con desvanecimiento lento.

48.6.2 Canales de trayectoria múltiple con desvanecimiento lento

Para las comunicaciones digitales por radioenlaces transhorizonte, se intenta mantener la linealidad de la transmisión; es decir, la salida del receptor debe ser una superposición lineal de la entrada del transmisor más el ruido del canal. Esto se logra operando el amplificador de potencia en la región lineal, o con amplificadores de potencia de saturación, por medio de técnicas de modulación de envolvente constante. Para los sistemas lineales, el desvanecimiento de trayectoria múltiple puede caracterizarse mediante una función de transferencia del canal $H(f; t)$. Esta función es el proceso aleatorio bidimensional en la frecuencia f y tiempo t que se observa como modulación de la portadora a la salida del canal cuando se aplica una excitación de onda sinusoidal a la frecuencia de la portadora a la entrada del canal. Para cualquier proceso aleatorio continuo es posible determinar la separación mínima requerida a fin de garantizar descorrelación con respecto a cada argumento.

Para la función de transferencia variante en el tiempo $H(f; t)$, sean t_d y f_d las separaciones de descorrelación en las variables de tiempo y frecuencia, respectivamente. Si t_d es una medida de la descorrelación del tiempo en segundos, entonces:

$$\sigma_t = \frac{1}{2\pi t_d} \quad \text{Hz}$$

es una medida de la velocidad de desvanecimiento o del ancho de banda del canal aleatorio. A menudo, la cantidad σ_t se denomina dispersión de Doppler, ya que es una medida del ancho del espectro recibido cuando se transmite una sola onda sinusoidal a través del canal. La relación dual para la descorrelación de frecuencia f_d en hertz sugiere que la variable de retardo:

$$\sigma_f = \frac{1}{2\pi f_d} \quad \text{s}$$

define la magnitud del retardo de trayectoria múltiple. La cantidad σ_f suele denominarse dispersión de retar-

dos de trayectoria múltiple, debido a que es una medida de la amplitud del proceso recibido en el dominio del tiempo cuando por el canal se transmite una sola función de pulso.

Algunos valores de estos factores de dispersión para las comunicaciones HF y por tropodispersión son:

HF	Tropodispersión
$\sigma_t \sim 0.1 \text{ Hz}$	$\sigma_t \sim 1 \text{ Hz}$
$\sigma_f \sim 10^{-3} \text{ s}$	$\sigma_f \sim 10^{-7} \text{ s}$

donde el símbolo \sim denota "del orden de".

Las dispersiones pueden definirse con precisión como momentos de espectros en un modelo de canal⁴, suponiendo estacionariedad en sentido amplio (WSS, de *wide-sense stationarity*) en la variable tiempo y dispersión no correlacionada (US, de *uncorrelated scattering*) en una variable retardo de trayectoria múltiple. Este modelo WSSUS y la suposición de estadísticos gaussianos para $H(f; t)$ proporcionan una descripción estadística en términos de una sola función correlación bidimensional del proceso aleatorio $H(f; t)$.

Esta caracterización ha sido bastante útil y exacta para una variedad de aplicaciones en enlaces por radio. Sin embargo, las suposiciones de estacionariedad y gaussiana no son necesarias para el uso de técnicas adaptativas de procesamiento de señales en estos canales. Lo que es necesario primero es que exista tiempo suficiente para "aprender" las características del canal antes de que cambien, y luego, que porciones descorrelacionadas de la banda de frecuencia sean excitadas de modo que sea posible descubrir un efecto de diversidad. Estas condiciones se reflejan en las dos siguientes relaciones en términos de los factores de canal ya definidos, la velocidad de transferencia de datos R y el ancho de banda B :

Restricción de aprendizaje

$$R(\text{b/s}) \gg \sigma_t \quad \text{Hz}$$

Restricción de diversidad

$$B(\text{Hz}) \gtrsim f_d \quad \text{Hz}$$

donde el símbolo \gtrsim significa "del orden de o mayor que".

La restricción de aprendizaje asegura que existe una razón de señal sobre ruido (SNR) suficiente para la comunicación confiable a velocidad R en el canal. Es claro que, si $R \sim \sigma_t$, entonces el canal cambiará antes de que sea posible reunir energía significativa para fines de medición. Cuando $R \gg \sigma_t$, los símbolos de los datos recibidos pueden considerarse el resultado de una señal de sondeo del canal, y el procesamiento idóneo puede generar estimaciones del carácter del canal durante esa época estacionaria específica. Las técnicas para el procesamiento de las señales en un receptor adaptativo no requieren necesariamente medir el canal directamente en la optimización del receptor, ya

que los requisitos de aprendizaje son aproximadamente los mismos. Si en la señal de sondeo solamente se utilizan símbolos de información, el modo de aprendizaje se denomina dirigido por decisiones. Cuando se utilizan símbolos digitales conocidos para el transmisor y para el receptor, el modo de aprendizaje se denomina dirigido por referencias. Una ventaja importante de los sistemas digitales es que, en muchas aplicaciones de comunicaciones adaptativas, la adaptación del receptor sin desperdicio de potencia para las señales de sondeo puede lograrse utilizando el modo dirigido por decisiones. Esto es posible en los sistemas digitales debido al número finito de parámetros o niveles en los símbolos transmitidos de la fuente y la elevada probabilidad de que las decisiones del receptor sean correctas.

La diversidad en aplicaciones con desvanecimiento se utiliza a fin de obtener canales de comunicaciones redundantes, de modo que, cuando algunos canales se desvanezcan, la comunicación siga siendo posible por los demás canales que no se encuentran en desvanecimiento.

En algunas de las formas de diversidad utilizada se controlan la separación mediante el empleo de antenas múltiples y el ángulo de llegada mediante el empleo de bocinas de alimentación, polarización, frecuencia y tiempo múltiples. Estas técnicas de diversidad algunas veces se denominan de diversidad explícita debido a su naturaleza externa visible.

Una forma alternativa de diversidad se denomina implícita debido a que el canal en sí proporciona redundancia. A fin de aprovechar esta diversidad implícita para mayor protección, es necesario utilizar técnicas de recepción con objeto de evaluar y combinar correctamente la información redundante. El potencial de la diversidad de frecuencia implícita se debe al hecho de que diferentes partes de la banda de frecuencia se desvanecen de forma independiente. Así, mientras una sección de la banda puede estar en desvanecimiento profundo, el resto puede utilizarse para comunicaciones confiables. Sin embargo, si el ancho de la banda transmitida B es pequeño en comparación con el intervalo de descorrelación de frecuencia f_d , entonces toda la banda se desvanecerá y no es posible obtener diversidad implícita. Por consiguiente, debe cumplirse la segunda restricción $B \geq f_d$ si se desea obtener una ganancia de diversidad implícita.

En sistemas con diversidad una pequeña descorrelación entre trayectorias alternativas de la señal puede proporcionar ganancia de diversidad importante. Así, no es necesario que $B \gg f_d$ para obtener ganancia de diversidad de frecuencia implícita, aun cuando la ganancia de diversidad implícita aumente claramente con la razón $B:f_d$. Obsérvese que la condición $R \ll B \geq f_d$ no excluye el empleo de la diversidad implícita, debido a que en el proceso de modulación es posible utilizar una técnica de expansión del ancho de banda a fin de distribuir la información transmitida sobre el ancho de banda disponible B . Se establecerá una diferencia entre estas condiciones de baja y alta velocidad de transferencia de datos, ya que las estructuras apropiadas del receptor asumen formas algo diferentes.

El efecto de la diversidad implícita descrito aquí resulta de la descorrelación en el dominio de la frecuencia en una aplicación de desvanecimiento lento ($R \gg \sigma_f$). Esta diversidad de frecuencia implícita puede en algunas circunstancias completarse por medio de un efecto de diversidad temporal implícito, que resulta de la descorrelación en el dominio del tiempo. En aplicaciones con desvanecimiento rápido ($R \geq \sigma_t$) es posible utilizar símbolos redundantes en un esquema de codificación a fin de obtener diversidad en el tiempo, suponiendo que la palabra codificada se extiende por más de una época de desvanecimiento. En la aplicación con desvanecimiento lento esta condición de extender la época del desvanecimiento puede lograrse intercalando las palabras del código a fin de obtener grandes intervalos de tiempo entre símbolos consecutivos en una palabra codificada específica. En el proceso de intercalación es necesario introducir un retardo de la señal más largo que la separación de la descorrelación temporal t_d . En muchas aplicaciones prácticas, en las que se requiere la transmisión de habla digitalizada, el retardo requerido es insatisfactoriamente grande para la comunicación del habla en dos sentidos. Por estas razones se hace más énfasis en las técnicas de diversidad de frecuencia implícita en los sistemas prácticos. Las estructuras del receptor que se abordarán a continuación pueden emplearse en situaciones en las que se aplique la diversidad de frecuencia implícita.

48.6.3 Estructuras de receptores adaptativos

Se considera un sistema de modulación de pulsos en amplitud en el que el conjunto muestra $\{a_k\}$ debe comunicarse por el canal utilizando una técnica de modulación que forme una correspondencia uno a uno entre la muestra a_k y la amplitud de un pulso transmitido. La modulación independiente de señales portadoras en cuadratura (es decir, $\sin 2\pi f_o t$ y $\cos 2\pi f_o t$, f_o = frecuencia portadora) se incluye en esta clase. Un ejemplo importante con las propiedades de detección óptimas es la manipulación por desplazamiento de fase en cuadratura (QPSK), que transmite el conjunto muestra $\{a_k = +1 + j\}$ cambiando el signo de los pulsos de la portadora en cuadratura según sea el signo de las partes real e imaginaria de la sucesión fuente $\{a_k\}$.

Receptores para canales con interferencia entre símbolos despreciable

Si cada muestra a_k puede ser una de M amplitudes posibles ($M = 4$ para la QPSK), entonces la velocidad de los datos transmitidos es:

$$R = [\log_2(M)]/T$$

donde $1/T$ es la velocidad de los símbolos transmitidos.

En la mayor parte de las aplicaciones terrestres de canales transhorizonte se utilizan medios limitados en la razón de señal sobre ruido, en vez de limitados en el ancho de banda. A fin de maximizar la detectabilidad de la señal, por lo general solamente se utilizan unas

cuantas amplitudes en estas aplicaciones. Cuando el periodo de símbolos T es mucho mayor que el ancho total de la dispersión de trayectoria múltiple del canal, sólo una pequeña parte de los símbolos adyacentes interfieren en la detección de un símbolo en particular. Para la aplicación con desvanecimiento lento, la restricción de la diversidad requiere que el ancho de banda de la señal B sea del orden de, o mayor que, el intervalo de descorrelación de frecuencia f_d . Entonces, las condiciones para una interferencia entre símbolos despreciable (ISI, de *intersymbol interference*) y del procesamiento adaptativo para obtener diversidad implícita son:

$$T \gg 2\pi\sigma_F = \frac{1}{f_d}$$

$$B \gtrsim f_d$$

Cuando el número de amplitudes M es pequeño, estas condiciones implican un sistema de baja velocidad de datos con respecto al ancho de banda disponible, es decir, un sistema de expansión del ancho de banda. La condición de baja velocidad de transferencia de datos para la diversidad de frecuencia implícita puede expresarse en términos de la velocidad de transferencia y el ancho de banda de los datos como:

$$R \ll B \log_2(M)$$

Si no hay interferencia entre símbolos, el esquema óptimo de detección contiene un filtro de ruido y un filtro acoplado al perfil de pulsos recibido⁵. El filtro óptimo de ruido tiene una función de transferencia igual al recíproco del espectro de la potencia del ruido $K(f)$. Cuando el ruido aditivo en la entrada del receptor es blanco, es decir, cuando su espectro es plano en la banda de frecuencia de interés, es posible omitir el filtro de ruido en el receptor óptimo. Así, el receptor óptimo para una función de transferencia de canal fija $H(f)$ contiene un filtro en cascada (en serie) cuyas funciones de transferencia componentes son:

$$R(f) = \frac{1}{K(f)} H^*(f)$$

donde el asterisco (*) representa conjugación compleja. En las diversas aplicaciones prácticas es necesario introducir un retraso de la señal a fin de hacer realizables estos filtros.

En general, K y H cambian con el tiempo y el receptor adaptativo debe seguir estas variaciones. El filtro de línea de retraso con derivaciones (TDL, de *tapped-delay-line*) es una estructura de filtro importante para tales aplicaciones de seguimiento de canal. El filtro TDL que se muestra en la figura 48-134 consta de una línea de retraso con derivaciones que multiplica la señal por el peso w_i de cada derivación. Para un sistema pasabanda con ancho de banda B , el teorema del muestreo establece que cualquier filtro lineal puede representarse mediante filtros TDL en paralelo que operen en cada componente de la portadora en cuadratura con separación entre derivaciones de $1/B$ o menos. Así, el receptor óptimo puede materializarse por medio de una cascada de estos dos filtros en cuadratura TDL en paralelo: uno con los pesos de las derivaciones ajustados a fin de formar el filtro de ruido, el segundo con dichos pesos ajustados para formar el filtro acoplado. Ya que la disposición en cascada de dos filtros lineales limitados en banda es otro filtro limitado en banda, en algunas aplicaciones es más conveniente utilizar un TDL a fin de obtener $R(f)$ directamente. En la práctica, las señales no pueden estar limitadas en tiempo y frecuencia a la vez, de modo que estos filtros TDL solamente pueden aproximar la solución ideal. Una ventaja del filtro TDL es la conveniencia de ajustar el voltaje de control del peso de las derivaciones como una forma de seguir las variaciones del canal y del espectro de ruido.

El diseño del receptor óptimo requiere un conocimiento del espectro de potencia del ruido $K(f)$ y de la función de transferencia del canal $H(f)$. Cuando $K(f)$ no es plana sobre la banda de interés, el proceso del ruido de entrada contiene una correlación que debe ser eliminada por el filtro de ruido. Algunas técnicas para reducir los efectos de ruido correlacionado son:

1. Predicción de valores futuros de ruido y cancelación de la componente correlacionada.
2. Técnicas de filtrado de error cuadrático medio utilizando un criterio de error apropiado.
3. Técnicas de excisión de ruido, empleando una transformada rápida de Fourier (FFT, de *fast Fourier transform*) para identificar y cortar picos de ruido en el dominio de la frecuencia.

El problema del filtrado de ruido suele ser importante en los sistemas de expansión de ancho de banda, debido a la interferencia de otros usuarios, así como a la amenaza de perturbaciones hostiles deliberadas.

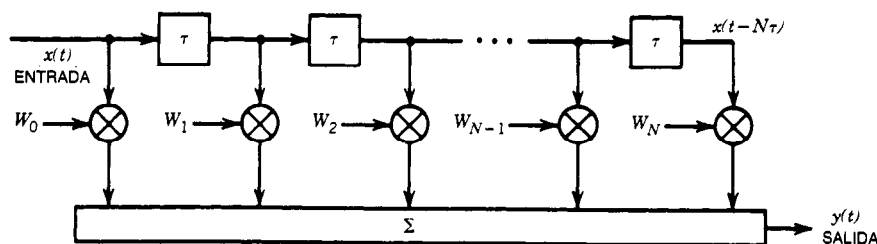


Fig. 48-134. Filtro de línea de retardo con derivaciones (TDL).

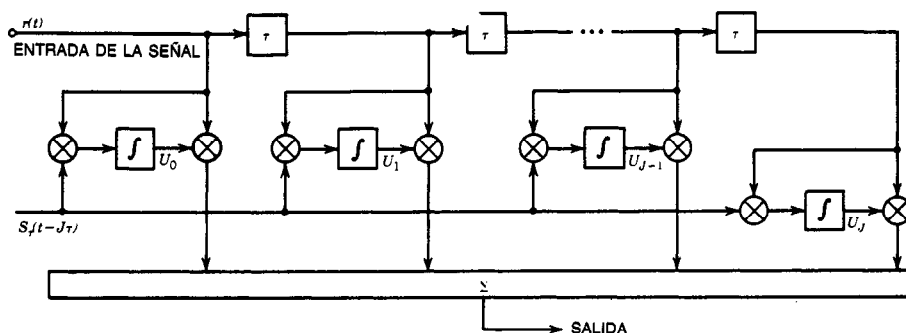


Fig. 48-135. Filtro RAKE.

Para materializar el filtro acoplado es posible utilizar un filtro TDL RAKE en el que se apliquen los conceptos desarrollados por Price y Green⁶, a fin de deducir de manera autoadaptable una aproximación a $H^*(f)$. El filtro RAKE recibe su nombre debido a que actúa como "rastrillo" para reunir todas las contribuciones de trayectoria múltiple. Esto puede lograrse utilizando el filtro TDL que se muestra en la figura 48-135, donde los pesos del TDL se deducen a partir de la correlación de los voltajes de derivación con una sucesión de prueba común, es decir, $S(t)$. Esta correlación da por resultado estimaciones de los valores de las derivaciones del canal TDL equivalente. Mediante el alineamiento en tiempo apropiado de la sucesión de prueba, los pesos del filtro RAKE se vuelven estimaciones de los valores de las derivaciones del canal, pero en orden temporal inverso, como se requiere en el diseño de un filtro acoplado. Para la adaptación del filtro RAKE, la sucesión de prueba puede ser una sucesión multicanalizada conocida con la información modulada, o decisiones del receptor utilizadas en una adaptación dirigida por decisiones.

Una estructura opcional para materializar el filtro acoplado es por medio de una línea de retardo recirculante que forme un promedio de los pulsos recibidos.

Esta estructura fue propuesta como una forma de reducir la complejidad en el diseño de un filtro RAKE⁷ para un sistema de manipulación por desplazamiento de frecuencia (FSK). Para un sistema de modulación de pulsos en amplitud (PAM, de *pulse amplitude modulation*), la estructura debe asumir la forma que se muestra en la figura 48-136. Una operación de modulación inversa entre la señal de entrada y la réplica local de la modulación de la señal se utiliza para quitar de la señal de llegada la modulación de la señal. Entonces, la línea de retardo recirculante puede formar un promedio del pulso recibido, que se utiliza en un correlacionador a fin de producir la salida del filtro acoplado. Este filtro de correlación es considerablemente más sencillo que el filtro TDL RAKE que se muestra en la figura 48-135.

Tanto en el filtro TDL RAKE como en el filtro de correlación acoplado se utiliza un proceso de promediación a fin de generar estimaciones del pulso de señal recibido.

Debido a que este pulso de señal se encuentra embebido en el ruido del receptor, en el proceso de medición debe obtenerse una razón de señal sobre ruido suficiente. Este requisito fundamental es la base de la restricción de aprendizaje:

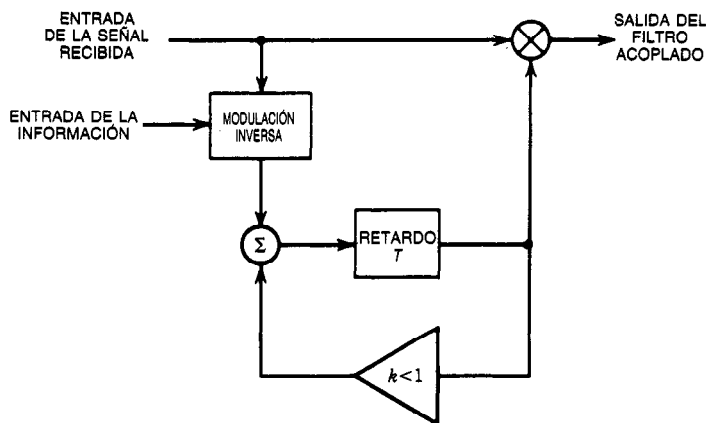


Fig. 48-136. Filtro de correlación para un sistema de modulación en amplitud de pulsos (PAM).

$$R(\text{b/s}) \gg \sigma_f \text{ Hz}$$

que ya se presentó. Si la velocidad de transferencia de la señal R con que se está efectuando la adaptación no es mucho mayor que la velocidad de cambio del canal σ_f , entonces el canal cambiará antes de que el procedimiento de promediación pueda acumular una razón de señal sobre ruido suficiente para realizar una medición exacta. Este requisito limita la aplicación de las técnicas de receptores adaptativos con ganancia de diversidad de frecuencia implícita a aplicaciones con desvanecimiento lento respecto a la velocidad de transferencia de datos. Afortunadamente, muchos canales tienen velocidades de desvanecimiento del orden de unos cuantos hertz, y los requisitos de datos, miles de veces mayores.

En general, el receptor para este ejemplo de ISI pequeña tiene un filtro de ruido a fin de acentuar las frecuencias en las cuales la potencia del ruido es más débil, así como una estructura de filtro acoplado que recombina coherentemente los elementos de la señal recibida a fin de obtener la ganancia de diversidad implícita. La diversidad implícita puede considerarse una diversidad de frecuencias debido a la descorrelación de las frecuencias recibidas. El filtro acoplado, en esta perspectiva, es un combinador de diversidad de frecuencias que combina coherentemente cada frecuencia conforme a su intensidad recibida. Sin el filtro acoplado podría ocurrir la combinación incoherente de las frecuencias recibidas y no presentarse efecto alguno de diversidad implícita.

Una aplicación importante de este sistema de baja velocidad de transferencia de datos se encuentra en ambientes de perturbación intencional, en los cuales se utiliza un exceso de ancho de banda a fin de reducir la vulnerabilidad a dicha perturbación. Sin embargo, en ambientes más benignos, la mayor parte de los requisitos de comunicación no permiten un gran ancho de banda respecto a la velocidad de transferencia de datos y, si en estas aplicaciones debe implementarse diversidad implícita, entonces es necesario reconsiderar el efecto de la interferencia entre símbolos.

Receptores de alta velocidad de transferencia de datos

Cuando la velocidad de transferencia de símbolos es del orden del intervalo de descorrelación de frecuencias del canal, las frecuencias en el pulso transmitido experimentarán diferentes variaciones de ganancia y fase, dando por resultado un pulso distorsionado.

Aunque en el transmisor puede no haber interferencia entre símbolos (ISI), la distorsión del pulso debida al medio del canal podría provocar interferencias entre muestras adyacentes de la señal recibida. En el dominio del tiempo, la ISI puede considerarse una "difuminación" del impulso transmitido debido a la trayectoria múltiple, provocando así superposición entre pulsos consecutivos. La condición es que la ISI pueda expresarse en el dominio de la frecuencia como:

$$T^{-1} \gg f_d \text{ Hz}$$

o, en términos de la dispersión de trayectoria múltiple:

$$T \lesssim 2\pi\sigma_f^{-1} \text{ s}$$

Dado que el ancho de banda de una señal PAM es del orden de por lo menos la velocidad de transferencia de datos T^{-1} hertz, no es necesaria la expansión del ancho de banda en condiciones ISI a fin de obtener ocupación por la señal de partes descorrelacionadas de la banda de frecuencia para diversidad implícita. Sin embargo, no resulta evidente si la presencia de interferencia entre símbolos puede destruir la ganancia de diversidad implícita disponible. Durante la década pasada se estableció que los receptores adaptativos pueden utilizarse para hacer frente a la interferencia entre símbolos y, en la mayor parte de los casos prácticos, concluir con una ganancia de diversidad implícita neta. Estas estructuras receptoras pueden clasificarse en tres clases: filtros de correlación con selección de tiempo, ecualizadores y detectores de máxima probabilidad.

Filtros de correlación. Estos filtros aproximan la parte del filtro acoplado del receptor óptimo sin ISI. El filtro de correlación que se muestra en la figura 48-136 podría no funcionar correctamente en presencia de ISI entre pulsos recibidos, debido a que el proceso de promediación podría sumar incoherentemente pulsos sobrepuestos. Cuando la dispersión de trayectoria múltiple es menor que el intervalo entre símbolos, esta condición puede aliviarse transmitiendo un pulso seleccionado en el tiempo, cuyo tiempo de "apagado" sea aproximadamente igual al ancho de la trayectoria múltiple del canal. La trayectoria múltiple provoca que el pulso transmitido seleccionado en el tiempo se difumine sobre toda la duración de los símbolos, aunque con poca o ninguna interferencia entre símbolos. Luego, es posible utilizar el filtro de correlación, a fin de acoplar el pulso recibido y proporcionar diversidad implícita⁸. En una configuración con diversidad tanto implícita como explícita es posible tolerar interferencia moderada entre símbolos debido a que la diversidad combinada suma coherentemente las componentes de la señal, e incoherentemente las componentes ISI. En virtud de que el tiempo de apagado del pulso no puede ser mayor del 100%, este método está, evidentemente, limitado en velocidad de transferencia de datos para condiciones de trayectoria múltiple fija. Además, la selección de tiempo en el transmisor da por resultado un aumento en ancho de banda, lo cual puede ser indeseable en una aplicación limitada en ancho de banda. La pérdida de potencia en los transmisores limitados en potencia pico, debida a la selección de tiempo, puede compensarse parcialmente empleando dos frecuencias portadoras con modulación de datos independiente⁹.

Ecualizadores autoadaptables. Los ecualizadores autoadaptables son sistemas de filtros lineales con parámetros ajustables electrónicamente que se controlan en un intento por compensar la interferencia entre símbolos. Los filtros de línea de retardo con derivaciones son una elección común para la estructura del ecualizador, debido a que los pesos de derivación constituyen un

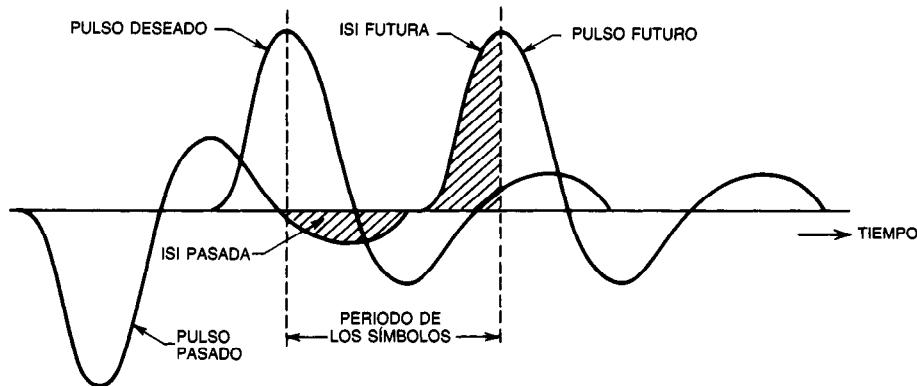


Fig. 48-137. Sucesión (secuencia) de pulsos recibidos después del filtrado del canal. ISI = interferencia entre símbolos.

conjunto conveniente de parámetros ajustables. Los ecualizadores autoadaptables se han utilizado ampliamente en aplicaciones de canales telefónicos¹⁰ para reducir los efectos ISI debidos al filtrado del canal. En una aplicación de canal de trayectoria múltiple con desvanecimiento, el ecualizador puede ejecutar tres funciones simultáneas: filtrado de ruido, filtrado acoplado para diversidad implícita y explícita, y eliminación de ISI. Estas funciones se efectúan adaptando un ecualizador de línea de retardo con derivaciones (TDLE, de *tapped-delay-line equalizer*) para obligar a que el error de medición sea mínimo. Diseñando el sistema de modo que la medición del error incluya la degradación debida a ruido correlacionado, ISI, filtrado y combinación inadecuada de diversidad, el TDLE minimizará sus efectos combinados.

Un **ecualizador lineal** (LE, de *linear equalizer*) se define como un ecualizador que filtra linealmente cada una de las N entradas de diversidad explícita. Se logra una mejora en los LE cuando se realiza un filtrado extra en las decisiones de los datos detectados. Debido a que utiliza decisiones en un esquema de retroalimentación, este ecualizador se conoce como ecualizador con retroalimentación de decisiones (DFE, de *decision-feedback equalizer*).

El funcionamiento del receptor de filtro acoplado de un LE y un DFE puede compararse a partir del análisis del ejemplo del tren de pulsos recibido de la figura 48-137. Los pulsos modulados binarios han sido difuminados por conducto del canal, produciendo la distorsión e interferencia de los pulsos desde los pulsos adyacentes. La detección ordinaria sin proyección de trayectoria múltiple integraría el proceso sobre un periodo de símbolo y decidiría la transmisión de $a + 1$ si el voltaje integrado es positivo y de $a - 1$ si el voltaje es negativo. La distorsión de los pulsos reduce de nuevo el margen en aquel proceso de integración. Un filtro acoplado correlaciona la onda de forma recibida con la réplica del pulso recibido, creando de esta forma el margen de ruido. La interferencia entre símbolos surge tanto de los pulsos futuros como pasados en estos sistemas de radio, ya que los contribuyentes de trayectoria múltiple cerca del retardo medio de trayec-

toria normalmente tendrán la mayor intensidad. Esta ISI puede compensarse en un ecualizador lineal empleando versiones desplazadas en el tiempo aproximadamente ponderadas de la señal recibida a fin de cancelar interferencias pasadas y futuras. En el DFE se utilizan versiones desplazadas en el tiempo de la señal recibida sólo para reducir la ISI futura. La ISI pasada se cancela por medio del filtrado de símbolos detectados pasados a fin de producir el voltaje de ISI correcto a partir de dichas interferencias. La propiedad de filtrado acoplado tanto en el LE como en el DFE se obtiene en la práctica mediante la separación de las derivaciones en el TDLE a intervalos más pequeños que el periodo de los símbolos.

En la figura 48-138 se muestra el DFE para un sistema de diversidad explícita de orden N . Se utiliza un TDLE de filtro directo (FF, de *forward filter*) para cada rama de diversidad a fin de reducir los efectos del ruido correlacionado, proporcionar filtrado acoplado y ponderación idónea para la combinación de diversidad explícita, y reducir los efectos ISI. Después de la combinación, desmodulación y detección de diversidad, las decisiones de los datos se filtran mediante un TDLE de filtro inverso (BF, de *backward filter*) a fin de eliminar la ISI de pulsos previos. Debido a que este filtro compensa estos efectos ISI "pasados", el filtro inverso solamente requiere compensar los ISI "futuros".

Se muestra un amplificador de control automático de ganancia (AGC, de *automatic gain control*) para cada rama de diversidad a fin de llevar la señal con desvanecimiento al intervalo dinámico del TDLE. También se muestra una señal de error dirigida por decisiones para la adaptación del DFE, como la diferencia entre la entrada y la salida del detector. Cualitativamente, es posible observar que si el DFE está bien adaptado, esta señal de error debería ser pequeña. La adaptación dirigida por referencias puede efectuarse multicanalizando un patrón de bits conocido dentro de la corriente del mensaje para su adaptación periódica.

Cuando se ignora el error de propagación debido a errores del detector, el DFE tiene el mismo o menor error cuadrático medio que el LE para todos los



Fig. 48-138. Ecualizador con retroalimentación de decisiones, diversidad de N -ésimo orden. AGC = control automático de ganancia; FF TDLE = ecualizador de línea de retardo con derivaciones de filtro directo; BF TDLE = ecualizador de línea de retardo con derivaciones de filtro inverso. DESM = desmodulación.

canales¹¹. El mecanismo de propagación de los errores se ha analizado por medio de cadenas de Markov¹² y se ha demostrado que es despreciable en aplicaciones prácticas de canal con desvanecimiento. Asimismo, en una aplicación de diversidad de orden N -ésimo, el número total de derivaciones TDLE suele ser menor para el DFE que para el LE. Esto se concluye debido a que en el primero solamente se utiliza un filtro inverso después de la combinación de los canales de diversidad en el filtro directo.

El rendimiento de un DFE en un canal con desvanecimiento puede predecirse¹³⁻¹⁵ utilizando una técnica de transformación que convierte la diversidad implícita en diversidad explícita y que trata los efectos ISI como una interferencia gaussiana. Como ejemplo, en la figura 48-139 se muestra la probabilidad media de error como función de la energía total recibida de los bits (E_b) respecto a la densidad espectral del ruido (N_0) para un sistema con diversidad cuádruple. La línea discontinua representa el rendimiento de dispersión de trayectoria múltiple cero ($\sigma_f = 0$), y las líneas continuas muestran el rendimiento para diferentes configuraciones de DFE (N = número de derivaciones del filtro directo y Δ = separación normalizada entre derivaciones) y condiciones ISI cuando la razón de dispersión de trayectoria múltiple entre el periodo de símbolos T es 0.25. Las condiciones sin ISI son cotas de rendimiento determinadas igualando a cero las componentes ISI. Cuando $\sigma/T = 0.25$, el funcionamiento debería estar a la derecha de la línea discontinua si no se utilizara el procesamiento de la señal adaptativa. Parece que el ecualizador elimina esta degradación y que también proporciona una ganancia de diversidad implícita, que es medida por la diferencia entre la línea continua $N = 3$, $\Delta = 0.5$ y la línea discontinua. La diferencia entre la curva $N = 3$, $\Delta = 0.5$ y la siguiente curva identificada como No ISI es la penalización de interferencia entre símbolos. Con los parámetros de filtro $N = 3$, $\Delta = 0.5$, ninguna técnica para eliminar la ISI puede producir mejores resultados

que esta curva. La pequeña penalización de ISI en este ejemplo típico es un argumento sólido para el uso del DFE en vez de técnicas de ISI más poderosas. Por último, la línea continua de la extrema izquierda se aproxima lo más que se puede, ya que los resultados muestran mejora despreciable a medida que se incrementa aún más el número de derivaciones. La pequeña diferencia observada muestra que un DFE con un número modesto de derivaciones de filtro directo se comporta casi tan bien como un DFE ideal con un número infinito de derivaciones.

Se ha desarrollado³ un modem DFE con velocidades de transferencia de datos hasta 12.5 Mb/s para aplicación en canales de tropodispersión con hasta cuatro órdenes de diversidad. En este modem DFE se utiliza solamente un TDLE con filtro directo de tres derivaciones y un TDLE con filtro inverso de tres derivaciones. Amplias pruebas en simuladores y en el campo^{3,13} han demostrado que se obtiene ganancia de diversidad implícita en un amplio intervalo de condiciones reales, a la vez que se eliminan casi por completo los efectos ISI. Por consiguiente, es posible el funcionamiento a velocidades de transferencia de datos cercanas a la distancia de descorrelación de frecuencia con pequeña penalización de interferencia entre símbolos. Los resultados medidos concuerdan con el funcionamiento predicho, del cual la figura 48-139 representa un ejemplo típico.

Detectores de máxima probabilidad. Ya que el DFE minimiza el voltaje de un detector analógico, es poco probable que sea óptimo para todos los canales con respecto a la probabilidad de errores en los bits. Considerando la interferencia entre símbolos como un código convencional definido en la recta real (o en la recta compleja para canales pasabanda), se han deducido^{16,17} algoritmos destinados a estimar sucesiones de máxima probabilidad para el canal PAM. Estos algoritmos constituyen un procedimiento de decodificación para decisiones del receptor que minimizan la

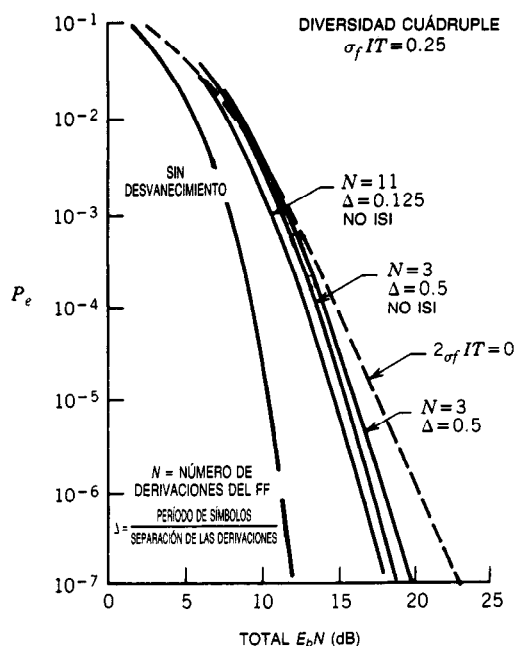


Fig. 48-139. Rendimiento de un ecualizador con retroalimentación de decisiones (DFE) de cuádruple diversidad. ISI = interferencia entre símbolos, FF = filtro directo. P_e = probabilidad de errores en los bits.

probabilidad de errores en la sucesión. Un receptor de estimación de sucesiones de máxima probabilidad (MLSE, de *maximum-likelihood sequence estimator*) sigue requiriendo un filtro de ruido y filtros acoplados para cada canal de diversidad. Después de estos filtros y operaciones combinadas, se utiliza una técnica de decodificación por enrejado a fin de determinar la su-

cesión transmitida de mayor probabilidad. En la figura 48-140 se ilustran gráficamente las funciones de filtrado, combinación y muestreo que preceden a la utilización del MLSE.

El algoritmo MLSE funciona asignando un estado a cada combinación de interferencia entre símbolos. Debido a la correspondencia uno a uno entre los estados y la ISI, es posible determinar la sucesión fuente de máxima probabilidad determinando la trayectoria de los estados.

Si se sabe que algún estado intermedio se encuentra en la trayectoria óptima, entonces la trayectoria de máxima probabilidad que empieza en ese estado y termina en el estado final será idéntica a la trayectoria óptima. Si en el instante n cada uno de los estados tiene asociada una trayectoria de máxima probabilidad que termine en ese estado, se concluye que en retrospectiva suficiente la historia de la trayectoria no dependerá del estado final específico al cual pertenece. La historia común de la trayectoria es la trayectoria de máxima probabilidad del estado¹⁶.

Dado que el número de combinaciones ISI —y por tanto el número de estados— es una función exponencial de la dispersión de trayectoria múltiple, el algoritmo MLSE presenta una complejidad que crece exponencialmente con dicha dispersión. Además, con ella aumenta linealmente el tamaño de estructura del ecualizador.

Asimismo, el requisito de combinación de diversidad y filtrado acoplado en el receptor MLSE impone aproximadamente la misma complejidad circuital y de implementación de adaptación que en un ecualizador para solamente este requisito.

Una comparación de las figuras 48-138 y 48-140 para el DFE y el receptor MLSE muestra que los sistemas son parecidos, excepto por la sustitución del filtro inverso en el DFE por el algoritmo de decodificación en el receptor MLSE. El filtro inverso es un filtro TDL

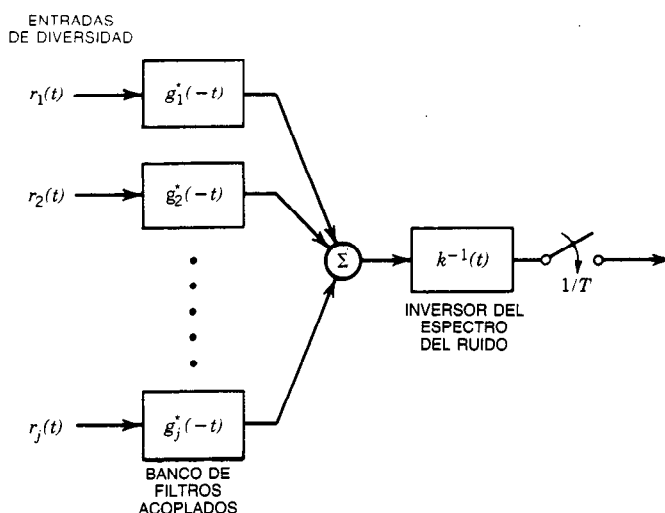


Fig. 48-140. Combinador de diversidad para un receptor de estimación de sucesiones (secuencias) de máxima probabilidad (MLSE).

con derivación en L , mientras que el algoritmo de decodificación del MLSE tiene complejidad computacional, que aumenta de manera exponencial con la dispersión de trayectoria múltiple. A cambio de esta mayor complejidad, el receptor MLSE tiene menor penalización (a veces nula) de interferencia entre símbolos para canales con desvanecimientos selectivos de frecuencia aislados y profundos. Sin embargo, en muchas aplicaciones en las cuales se utilizan órdenes elevados de diversidad, estos desvanecimientos selectivos de frecuencia profundos no ocurren tan a menudo como para afectar significativamente a la probabilidad media de error. Este resultado se ilustra en la curva de rendimiento que se proporciona en la figura 48-139, que muestra solamente una pequeña penalización ISI para el DFE con sólo tres derivaciones, en comparación con el DFE con igual número de derivaciones cuando no existe ISI.

48.6.4 Nuevas áreas de investigación

En los ecualizadores adaptativos actuales para aplicaciones en canales con desvanecimiento se utiliza un algoritmo de gradiente estimado para el seguimiento del canal que puede ser bastante lento para canales con desvanecimientos selectivos de frecuencia profundos o en presencia de interferencia electrónica hostil. Los algoritmos deducidos a partir de las ecuaciones de estimación de Kalman han sido sugeridos¹⁸ como medios para hacer realidad todo el potencial de la capacidad de seguimiento adaptativo.

Un seguimiento más rápido podría proporcionar ímpetu a la ecualización HF, donde las velocidades de transferencia de datos digitales no siempre son muchos órdenes de magnitud mayores que la velocidad de cambio del canal. Los receptores adaptativos en los que se utilicen antenas múltiples en un ambiente de canal de trayectoria múltiple con desvanecimiento pueden proteger contra perturbaciones que anulen la recepción en la antena, además de proporcionar ganancia de diversidad implícita. Los algoritmos de seguimiento más rápidos podrían incrementar la flexibilidad del sistema a un intervalo más amplio de amenazas de perturbación deliberada.

REFERENCIAS BIBLIOGRÁFICAS

1. V. I. Tatarski, "The Effects of the Turbulent Atmosphere on Wave Propagation", Israel Program for Scientific Translation, Jerusalén, Israel.
2. P. Monsen, "High Speed Digital Communication Receiver", U.S. Patent 3 879 664.
3. D. R. Kern y P. Monsen, "Megabit Digital Troposcatter Subsystem (MDTS)", GTE Sylvania, Needham, MA, y SIGNATRON, Lexington, MA, Informe final ECOM-74-0040-F.
4. P. A. Bello, "Characterization of Randomly Time-Variant Linear Channels", *IEEE Trans. Commun. Syst.* **CS-11**:360-393.
5. L. A. Wainstein y V. D. Zubakov, *Extraction of Signals from Noise*, Prentice-Hall, Englewood Cliffs, NJ.
6. R. Price y P. E. Green, Jr., "A Communication Technique for Multipath Channels", *Proc. IRE* **46**:555-569.
7. S. M. Sussman, "A Matched Filter Communication System for Multipath Channels", *IRE Trans. Inform. Theory* **IT-6**:367-372.
8. M. Unkauf y O. A. Tagliaferri, "An Adaptive Matched Filter Modem for Digital Troposcatter", *Conf. Rec., Int. Conf. Commun.*
9. M. Unkauf y O. A. Tagliaferri, "Tactical Digital Troposcatter Systems", en *Conf. Rec., Nat. Telecommun. Conf.*, vol. 2, pp. 17.4.1-17.4.5.
10. R. W. Lucky, J. Salz y E. J. Weldon, Jr., *Principles of Data Communication*, McGraw-Hill, Nueva York.
11. P. Monsen, "Feedback Equalization for Fading Dispersive Channels", *IEEE Trans. Inform. Theory* **IT-17**:56-64.
12. P. Monsen, "Adaptive Equalization of the Slow Fading Channel", *IEEE Trans. Commun.* **COM-22**.
13. P. Monsen, "Theoretical and Measured Performance of a DFE Modem on a Fading Multipath Channel", *IEEE Trans. Commun.* **COM-25**:1144-1153.
14. M. Schwartz, W. R. Bennett y S. Stein, *Communications Systems and Techniques*, McGraw-Hill, Nueva York.
15. P. Monsen, "Digital Transmission Performance on Fading Dispersive Diversity Channels", *IEEE Trans. Commun.* **COM-21**:33-39.
16. G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", *IEEE Trans. Inform. Theory* **IT-18**:363-377.
17. G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data Transmission Systems", *IEEE Trans. Commun.* **COM-22**:624-636.
18. D. Godard, "Channel Equalization Using a Kalman Filter for Fast Data Transmission", *IBM J. Res. Develop.*, pp. 267-273.

48.7 ANÁLISIS COMPARATIVO DE LAS TÉCNICAS DE COMUNICACIONES

E. T. Dickerson

En las secciones 47.2, 47.3 y 48.1 se han presentado varias técnicas de comunicaciones. Se analizaron dos clases de sistemas de comunicaciones: las técnicas analógicas, en las cuales los procesos de modulación son funciones continuas del tiempo, y las técnicas digitales, en las que los procesos de modulación se encuentran en forma de sucesión de símbolos de datos discretos.

En esta sección se comparan varias técnicas de modulación que pertenecen a estas dos grandes clases. Durante el análisis comparativo podrá observarse que se requieren diferentes criterios de comparación para las técnicas de modulación analógica y digital.

48.7.1 Criterios de comparación

La selección del "mejor" diseño de sistema de comunicaciones para una aplicación específica requiere la definición del término "mejor" para tal aplicación. Existen varios criterios ampliamente aceptados para evaluar cuantitativamente el rendimiento de un sistema de comunicaciones. Algunos de ellos son:

1. El ancho de banda requerido.
2. La probabilidad de error (en un sistema digital).
3. La razón de señal recibida sobre ruido (en un sistema analógico).

Otras medidas de calidad pueden aplicarse a un caso particular *conforme a la definición del usuario*. Es decir, lo que es mejor en un caso puede no serlo en otro. La evaluación del rendimiento está restringida por condiciones que se relacionan con la *idoneidad* de un sistema candidato para una aplicación específica. Las restricciones pueden incluir una o más de las siguientes:

1. Complejidad.
2. Costo.
3. Potencia disponible.
4. Habilidad del operador.
5. Nivel tecnológico aplicado.

Por supuesto, la lista no es exhaustiva, sino que ilustra las restricciones prácticas que deben considerarse. Es evidente que no existe un mejor sistema de comunicaciones, ya que de ser así no habría tal amplia variedad para elegir.

Analógico frente a digital

Una división importante de los sistemas de comunicaciones que se utilizará en esta sección es: sistemas de modulación de ondas de forma continuas (analógicas) y sistemas de modulación de ondas de forma discretas (digitales). No existe una elección sencilla y general entre la implementación analógica y la digital. Para algunas aplicaciones, tales como la comunicación de datos de computadora ya almacenados en forma digital, claramente el sistema digital parece ser el más indicado. Por otra parte, la transmisión por televisión de calidad demanda un sistema analógico, conforme a la mayor parte de los criterios razonables. Por tanto, el método de evaluación que se adopta en el análisis siguiente es considerar por separado los sistemas de comunicaciones analógico y digital. A fin de evitar planteamientos prolongados, se utilizarán técnicas tabulares y gráficas.

48.7.2 Comparación de técnicas de ondas de forma continua (analógica)

Aquí se consideran dos categorías de técnicas de modulación analógicas, a saber, la modulación lineal y la modulación angular o exponencial (modulación no lineal). A fin de comparar varios esquemas de modulación lineal y no lineal, es necesario definir el criterio

según el cual habrán de juzgarse los sistemas. La relación entre la razón de señal sobre ruido antes y después de la detección (SNR_i y SNR_o , respectivamente) se utiliza aquí como una medida de "bondad" al comparar las distintas técnicas de modulación analógica. A fin de establecer matemáticamente esta relación se requiere un breve análisis de la teoría de la información.

La teoría de la información, por supuesto, se aplica tanto a los sistemas de comunicaciones digitales como a los analógicos. Sin embargo, la aplicación a estos últimos sistemas no es tan evidente como a los primeros (sección 47.3, Teoría de la información). Esto se debe a que los resultados más importantes de la teoría de la información se expresan en términos de mediciones discretas (bits y b/s). Estos principios y mediciones pueden aplicarse a sistemas de comunicaciones analógicos considerando la representación directa equivalente de la información analógica.

El teorema de muestreo de Nyquist garantiza la reproducción exacta de un proceso analógico muestreado limitado en banda, si las muestras se toman a una frecuencia igual a dos veces la mayor frecuencia contenida en la señal. Si el número de niveles de cuantización por muestra se elige de modo que represente con suficiente precisión la señal, entonces el producto de muestras/segundo y bits/muestra representa con aproximación aceptable la velocidad de transferencia de información en bits/segundo. Por ejemplo, si una señal de habla está limitada en banda a 4 kHz, se muestrea a 8 ks/s y se cuantiza a 64 niveles (6 b/s), entonces la velocidad aproximada de transferencia de datos es 48 kb/s. Con base en una aplicación rigurosa de la teoría de la información, la velocidad de transferencia de datos real puede ser menor que la velocidad de transferencia de símbolos binarios, debido a que la medida de información depende de la distribución de probabilidad de la señal. Sin embargo, desde el punto de vista del diseñador, la velocidad de transferencia de símbolos binarios es un parámetro significativo prácticamente equivalente a la velocidad de transferencia de información.

Según la ley de Shannon-Hartley, la razón de señal sobre ruido en el ancho de banda de predetección, SNR_i , determina la velocidad máxima a la que puede llegar la información al detector. Matemáticamente, esta capacidad del canal C_i está dada por:

$$C_i = B_i \cdot \log_2[1 + (SNR)_i]$$

donde C_i = velocidad máxima de transferencia de información en la entrada del sistema, b/s

B_i = ancho de banda de predetección, Hz

$(SNR)_i$ = razón de potencia de la señal entre potencia del ruido en el ancho de banda B_i

En un sistema de comunicaciones ideal, la velocidad de transferencia de información es constante a través de todo el sistema. En particular, dicha velocidad es igual a la entrada y a la salida del receptor, lo que implica una pérdida de información despreciable debida a errores, y a su vez implica $SNR_i \gg 1$. La definición de "ideal" puede ampliarse aún más a fin de estipular que la velocidad de transferencia de información

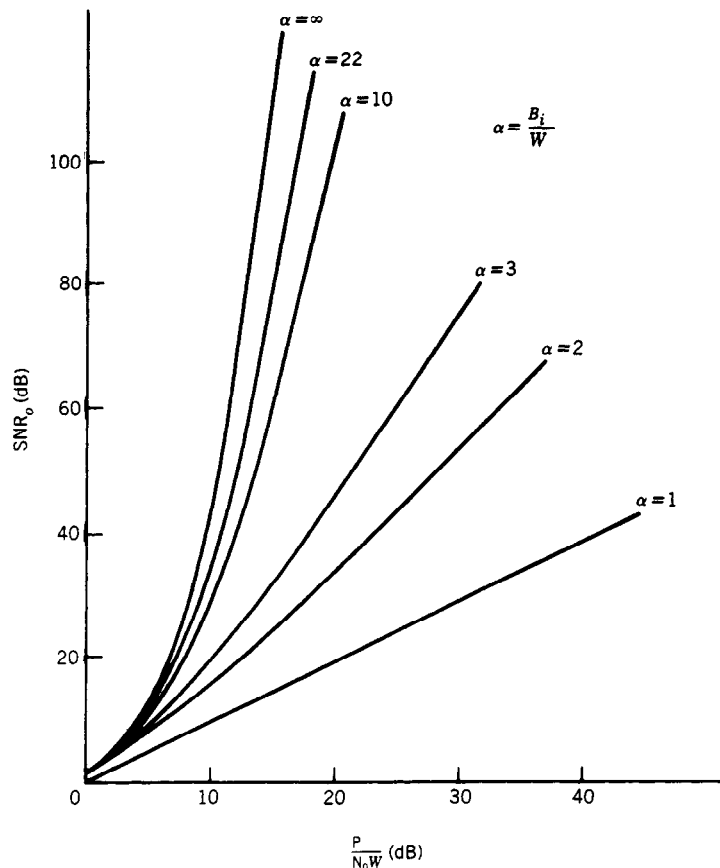


Fig. 48-141. Curvas de rendimiento para moduladores y detectores "ideales" en función del factor de expansión del ancho de banda.

es igual a la capacidad del canal definida antes. En estas condiciones, se ha demostrado (Panter, pág. 605) que las razones de señal sobre ruido de entrada y de salida están relacionadas exponencialmente de la siguiente forma:

$$\text{SNR}_o \cong (\text{SNR}_i)^\alpha$$

o bien:

$$\text{SNR}_o \cong (P/\alpha N_o W)^\alpha$$

donde P = potencia de señal a la entrada del receptor, W

N_o = densidad espectral de potencia de ruido unilateral, W/Hz

C_o = velocidad máxima de transferencia de información en la salida, b/s

W = ancho de banda de la señal, Hz

α = factor de expansión del ancho de banda, $\alpha W = B_i$

En la figura 48-141 se presentan gráficas del SNR_o de posdetección en función de $P/N_o W$ para diferentes valores del factor de expansión del ancho de banda.

Para valores de α mayores que 22 se obtiene poca mejora con el desarrollo ulterior del ancho de banda de transmisión B_i .

Ahora se analizará el rendimiento de algunos sistemas prácticos de modulación-desmodulación analógica en presencia de ruido blanco gaussiano aditivo. La figura 48-142 es también una gráfica de SNR_o como función de $P/N_o W$. En esta figura se incluyen curvas de ganancia de detección para diversos esquemas de modulación analógica. La detección ideal de modulación de amplitud (AM), de banda lateral doble (DSB) y de modulación DSB en cuadratura daría por resultado una curva de ganancia de detección correspondiente a un factor de expansión del ancho de banda igual a 2. Sin embargo, como se muestra, la detección coherente de estas ondas de forma se encuentran lejos del caso ideal. También se muestra la curva de ganancia de detección para FM, con un factor de expansión del ancho de banda igual a 22. Aunque el sistema FM no es ideal, se utiliza ampliamente y es muy flexible. El efecto de umbral para el sistema FM está bien definido. Las curvas también indican que el funcionamiento FM para razones de señal sobre ruido menores de 10 dB es marginal debido al efecto de umbral.

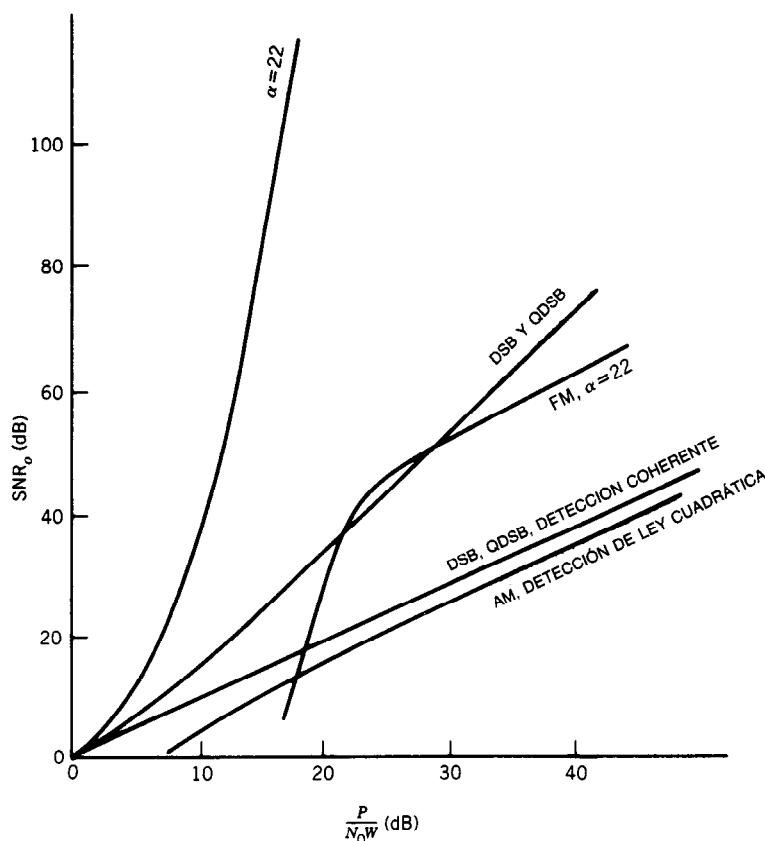


Fig. 48-142. Comparación de esquemas analógicos realizables con el esquema de modulación "ideal". DSB = doble banda lateral; QDSB = DSB en cuadratura; AM = modulación de amplitud.

Para completar, en la tabla 48-20 se hace una comparación relativa de varias técnicas de modulación analógica lineal y no lineal. En el análisis de las técnicas de modulación lineal que se presenta aquí, se supone que la componente de ruido de la señal recibida es ruido gaussiano blanco aditivo con densidad espectral de potencia de N_0 watts por hertz. La SNR_i puede expresarse como la razón de la potencia de la señal P a la entrada del receptor y la potencia de ruido $N = N_0B_i$ en el ancho de banda de predetección B_i . En la tabla 48-20 se utilizan como criterios de comparación la razón de señal sobre ruido SNR_i de entrada y la razón de señal sobre ruido después de la detección de la señal SNR_o .

Como se indica en la tabla, las técnicas lineales de supresión de la portadora, de banda lateral única (SSB), de banda lateral doble (DSB) y de banda lateral vestigial (VSB) son "mejores" que la AM cuando se juzgan con base en la ganancia de detección (SNR_o/ SNR_i). Además, la SSB y la DSB no están sujetas al efecto de umbral que degrada el rendimiento de la AM cuando se utiliza detección de envolvente. El amplio intervalo de técnicas de modulación analizado aquí permite que grupos de usuarios con diversos requisitos

obtengan un sistema de comunicaciones que sea el mejor para sus necesidades. Por ejemplo, la FCC (Federal Communications Commission) limita la potencia del transmisor a 1 000 W promedio y a 2 000 W pico en el caso de la radio para aficionados. La ocupancia espectral también se restringe bastante para este tipo de radio, a varias bandas estrechas esparcidas por todo el espectro de la radio. Por todas estas razones, los radioaficionados desean pagar el precio de una mayor complejidad del equipo a cambio de obtener ventajas de ancho de banda y potencia características del sistema SSB. Otra ventaja del SSB es que se eliminan los heterodinos de portadora (interferencia de tono) inherentes al sistema AM.

El sistema AM es atractivo para un grupo de usuarios muy diferente. Este grupo requiere la transmisión a un mayor número de escuchas a un costo de equipo mínimo. Estos requisitos se cumplen con un sistema AM en el que se utilice un detector de envolvente.

Las técnicas analógicas lineales analizadas hasta el momento requieren una amplificación lineal después de efectuado el proceso de modulación. A menudo es deseable utilizar técnicas eficientes para la amplificación de potencia. Por ejemplo, en una gran cantidad de

Tabla 48-20. Comparación de los esquemas de modulación analógica

Tipo de modulación	Ancho de banda de transmisión ^a	Ganancia de detección SNR_d/SNR_i	Factor de expansión del ancho de banda, B/W^b	Respuesta de CD	Complejidad del equipo	Aplicaciones populares
De doble banda lateral (DSB)	W	1	2	Sí	Moderada: se requiere demodulación coherente; el receptor puede simplificarse mediante la transmisión de un tono piloto	Datos analógicos
De amplitud (AM)	$2W$	$1/3$	2	No	Secundaria: se utiliza detección de envolvente simple; el modulador es sencillo; en el receptor se requiere aislamiento de CD	Radiodifusión
De banda lateral única (SSB)	$2W$	1	1	No	Importante: se requiere desmodulación coherente; el modulador requiere filtrado complicado de la banda lateral	De punto a punto, comunicación del habla, radio para aficionados
De banda lateral estigial (VSB)	$2W$	1	2	Sí	Importante: se requiere desmodulación coherente; se requieren filtros simétricos complicados	Digital y analógica, transmisión de banda ancha
De VSB más portadora (VSB + C)	$2W$	$1/3$	2	No	Moderada: se utiliza detector de la envolvente; se requiere filtrado simétrico	Señales de banda ancha, vídeo para transmisiones de TV
De frecuencia (FM)	$2(D + 1)W$	$3/2D^2$	$2(D + 1)$	Sí	Moderada: transmisor complicado; en el receptor se utiliza un discriminador de frecuencias sencillo	Radiodifusión, de punto a punto, relevador de microondas, audio para TV
De fase (PM)	$2(D + 1)W$	K_p^2	$2(D + 1)$	Sí	Moderada: se requiere una simple modificación del receptor de FM	Habla para transmisión de datos
De la banda base	W	1	1	Sí	Secundaria: no se requiere modulación o traslación de frecuencia	Comunicaciones entre computadoras, comunicaciones de punto a punto cortas

^a W = ancho de la banda base de la señal; D = constante de desviación.^b B = ancho de banda de transmisión

aplicaciones espaciales se utiliza un amplificador de potencia de tubo de ondas progresivas (TWT, de *traveling-wave tube*) que debe funcionar en el modo saturado (no lineal) a fin de obtener una operación óptima. Esta aplicación requiere que la señal modulada tenga una envolvente constante y que la información esté contenida en el ángulo de fase. De todas las técnicas mencionadas hasta el momento, la FM y la PM satisfacen este requisito. En comparación con otras técnicas de modulación analógica, la FM proporciona la mejor ganancia de detección. Este funcionamiento mejorado se logra a expensas del ancho de banda.

Como se indica en la tabla 48-20, el sistema FM ofrece considerables ahorros en potencia para un funcionamiento equivalente, medido en términos de la ganancia de detección, sobre los otros esquemas analógicos analizados hasta el momento. Los esquemas VSB y SSB son los más complicados, mientras que el esquema AM es el más sencillo de realizar en términos de equipo. Este análisis de los esquemas de modulación analógica ha sido breve, por lo que solamente se han contemplado las técnicas fundamentales. Varios sistemas de comunicaciones requieren una combinación de estas técnicas básicas. Para esta señal compleja se utilizan moduladores de banda lateral vestigial más portadora (VSB + C), de frecuencia (FM) y de doble banda lateral en cuadratura (QDSB), y multicanalizadores por división de frecuencia (FDM) y por división de tiempo (TDM), respectivamente.

48.7.3 Comparación de técnicas de ondas de forma discreta (digital)

Probabilidad de error

Como se indicó en la sección 48.7.1, el criterio más importante para la comparación de las técnicas de modulación digital es la probabilidad de que el receptor cometa un error al detectar la señal recibida en presencia de ruido gaussiano blanco aditivo. En general, la probabilidad de una decisión incorrecta se considera como la probabilidad de un error de un símbolo. Para el sistema binario, un símbolo suele contener 1 bit de información.

Para facilitar la comparación, inicialmente sólo se considerarán técnicas de modulación binaria; es decir, los casos en los cuales la probabilidad de errores en los símbolos y la probabilidad de errores en los bits son iguales. Sin embargo, para hacer más completa esta exposición aquí se presenta la relación entre las proporciones de errores en los bits binarios y las proporciones de errores en los bits M -arios. Para el caso de señalización ortogonal:

$$P_E(\text{binario}) = P_E(M\text{-ario}) \cdot [2^{k-1}/(2^k - 1)]$$

donde $P_E(\text{binario})$ es la probabilidad de error para un esquema de señalización binario y $P_E(M\text{-ario})$ es la probabilidad de error para esquemas de señalización M -arios.

Para señales no ortogonales, suponiendo sólo un bit por símbolo, las probabilidades de error están relacionadas por:

$$P_E(M\text{-ario}) = P_E(\text{binario}) \log_2 M = P_E(\text{binario}) \cdot k$$

donde M es igual a 2^k bits por símbolo.

Ya que las señales consideradas aquí son deterministas y el ruido aditivo es un proceso aleatorio con distribución gaussiana, la señal más el ruido presentarán una función densidad de probabilidad con distribución de Rice para razones de señal sobre ruido grandes.

Suponiendo que la señal recibida es una sucesión independiente equiprobable de bits y que se utiliza un receptor de correlación, la probabilidad de error P_E para una señal binaria manipulada por desplazamiento de fase (PSK) puede expresarse como:

$$P_E = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{\exp(-u^2/2)}{\sqrt{(E_b/N_0)(1 - \cos \theta)}} du$$

donde E_b = energía de la señal por bit, J

N_0 = densidad espectral de potencia de ruido unilateral, W/Hz

θ = ángulo entre las ondas de forma de señalización en el espacio de señales

La razón E_b/N_0 se utiliza para describir el rendimiento del sistema para diversas técnicas de modulación digital. Es fácil relacionar esta cantidad con la razón de señal sobre ruido en el ancho de banda de la velocidad de transferencia de bits en el receptor, siempre que se conozcan la velocidad de señalización R y el ancho de banda de la velocidad de transferencia de bits W . La siguiente ecuación describe esta relación:

$$P/N_0W = (E_b/N_0)(R/W)$$

En la figura 48-143 se comparan varios esquemas de modulación con base en la probabilidad de un error de un bit. Estas gráficas de probabilidad de error contra E_b/N_0 indican que los esquemas coherentes suelen proporcionar mejores resultados que los esquemas incoherentes, y que la PSK suele funcionar mejor que la FSK. Por ejemplo, para una probabilidad de error igual a 10^{-4} , la FSK incoherente requiere que E_b/N_0 sea aproximadamente 4 dB mayor que la probabilidad requerida para la PSK coherente. Así, la solución intermedia entre funcionamiento y complejidad es obvia. En una situación limitada en potencia, como se encuentra en sistemas espaciales, la justificación del mayor costo debido al aumento en la complejidad del sistema es evidente. Como lo señala Sklar, cuando se elige un sistema de comunicaciones existen otras consideraciones además de P_E y la complejidad del sistema. Por ejemplo, en algunos casos (como en condiciones de propagación con desvanecimiento aleatorio), un sistema incoherente es más robusto y deseable debido a que puede haber dificultades para establecer una referencia portadora para los esquemas coherentes.

Plano ancho de banda-eficiencia

Este plano es una representación gráfica eficaz para comparar varias técnicas de señalización digital. La figura 48-144 y el análisis de los parámetros se han adaptado de Sklar. El teorema de Shannon-Hartley puede

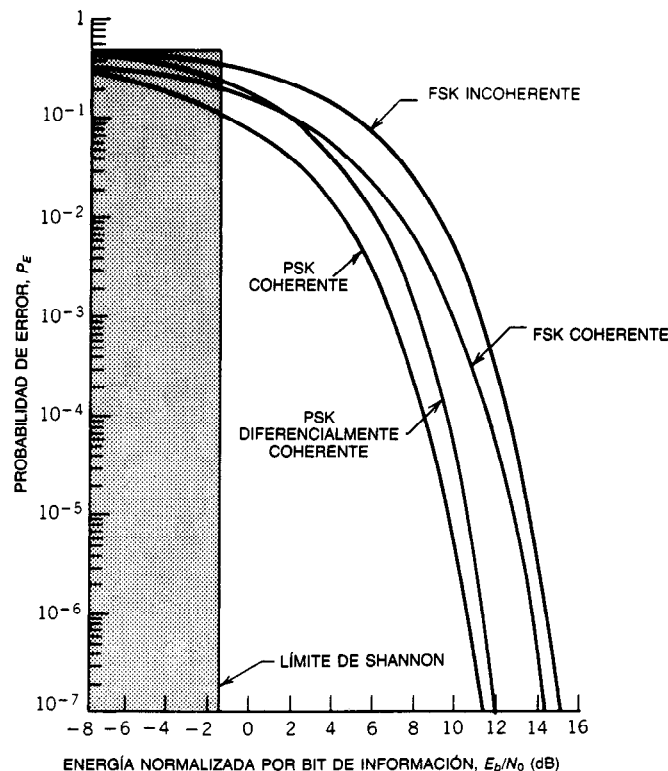


Fig. 48-143. Probabilidad de errores en los bits para técnicas de modulación binaria. FSK = manipulación por desplazamiento de frecuencias; PSK = manipulación por desplazamiento de fase.

escribirse matemáticamente en términos de E_b/N_0 y de W/R , suponiendo que la velocidad de la información de entrada y de salida del receptor es la misma. La siguiente ecuación se dedujo utilizando esta ley:

$$E_b/N_0 = W/R(2^{R/W} - 1)$$

En la figura 48-144 se ha graficado esta ecuación en el plano R/W en función de E_b/N_0 . La ordenada R/W es una medida de cuántos datos es posible transmitir en un ancho de banda específico en un tiempo dado; por tanto, refleja con qué eficiencia se utiliza el recurso del ancho de banda. La abscisa es la métrica E_b/N_0 en decibels. Si la capacidad de canal C y la velocidad de señalización R son iguales en la ecuación anterior, la curva discontinua representa una frontera que separa la comunicación potencial libre de errores de las regiones en las que no es posible la comunicación. En el diagrama ancho de banda-eficiencia de la figura 48-144 se encuentran graficados puntos para modulación PSK mínima (MPSK) y FSK mínima (MFSK), con $P_E = 10^{-5}$. Para la modulación MPSK, R/W aumenta con M creciente; sin embargo, para la modulación MFSK, R/W disminuye con M creciente. La ubicación de los puntos MPSK indica que la PSK binaria (BPSK) ($M = 2$) y la PSK en cuadratura (QPSK) ($M = 4$) requieren el mismo E_b/N_0 . Es decir, para el mismo valor

de E_b/N_0 , la QPSK tiene una eficiencia de ancho de banda de 2 b/s/Hz, en comparación con el valor de 1 b/s/Hz de la BPSK. Esta característica única surge del hecho de que la QPSK es efectivamente una composición de dos señales BPSK, transmitidas en formas de onda ortogonales entre sí y que presentan la misma ocupancia espectral. En el plano ancho de banda-eficiencia también están graficados los puntos de operación para la modulación MFSK incoherente con $P_E = 10^{-5}$. La posición de los puntos MFSK indica que la FSK binaria (BFSK) ($M = 2$) y la FSK en cuadratura (QFSK) ($M = 4$) tienen la misma eficiencia de ancho de banda, aun cuando la primera requiere mayor E_b/N_0 para la misma proporción de errores. La eficiencia de ancho de banda varía con el índice de modulación.

La frontera teórica de capacidad de información se observa en el plano ancho de banda-eficiencia. Esta frontera es el límite teórico de la velocidad a la cual es posible transmitir la información libre de errores a través del canal. Esta velocidad representa una meta para el diseñador cuando se especifica el rendimiento deseado. También se muestran curvas con igual probabilidad de error para diversos esquemas de modulación y desmodulación. Las curvas identificadas por P_{B1} , P_{B2} y P_{B3} son construcciones hipotéticas de algún esquema arbitrario de modulación y codificación; la curva P_{B1}

representa la mayor probabilidad de error de las tres curvas y la curva P_{B3} representa la menor. En la figura se muestra la dirección general en la que se mueven las curvas para mejorar.

Es posible concebir soluciones intermedias entre P_E y E_b/N_0 utilizando el diagrama ancho de banda-eficiencia. Tales soluciones son cambios en el punto de operación en la dirección que se indica con las flechas. El movimiento del punto de operación a lo largo de la línea 1 puede considerarse un compromiso en el funcionamiento de P_E en función de E_b/N_0 , con R/W fijo. De forma parecida, el movimiento a lo largo de la línea 2 se considera un compromiso en el funcionamiento de P_B en función de W (o de R/W), con R/W fijo. Por último, el movimiento a lo largo de la línea 3 ilustra el

compromiso de W (o R/W) en función de E_b/N_0 , con P_E fijo. El movimiento a lo largo de la línea 1 se efectúa incrementando o reduciendo el cociente E_b/N_0 disponible. El movimiento a lo largo de la línea 2 o de la línea 3 se realiza modificando el esquema de modulación o codificación.

El rendimiento del sistema se ha descrito para varios esquemas de señalización digital. Se ha evaluado mediante la probabilidad de error en función de parámetros tales como la densidad espectral de potencia del ruido, velocidad de señalización y potencia de la señal. La métrica más importante utilizada fue el cociente E_b/N_0 , relacionado con los tres parámetros del sistema que se acaban de mencionar. En la tabla 48-21, que es una adaptación de Shanmugam, se resumen los esque-

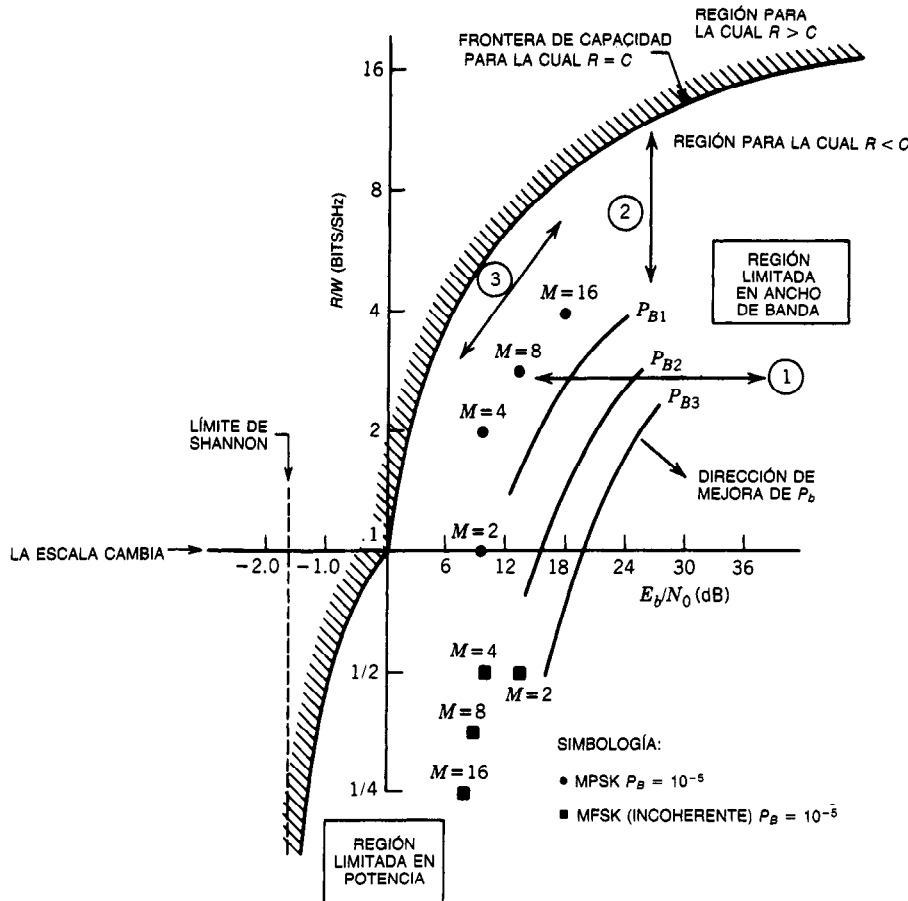


Fig. 48-144. Diagrama ancho de banda-eficiencia.

$$R/W \leq C/W = \log_2(1 + S/N) = \log_2[1 + E_b/N_0(C/W)]$$

Cualquier esquema de señalización digital que transmita $\log_2 M$ bits en T segundos, utilizando un ancho de banda W (Hz), funciona siempre a una eficiencia de $R/W = \log_2 M/WT$ (b/s/Hz). Frontera de la capacidad: lugar geométrico de mínimo E_b/N_0 en función de R/W para transmisión confiable. Límite de Shannon: límite de $E_b/N_0 = -1.6$ dB cuando C/W tiende a cero. MPSK = manipulación por desplazamiento de fase múltiple, MFSK = manipulación por desplazamiento de frecuencia múltiple. Tomado de Sklar, reproducido con permiso.

mas de modulación digital descritos en esta sección. A continuación se consideran los parámetros utilizados para el análisis comparativo en esta tabla.

Ancho de banda. Los requisitos de ancho de banda para un sistema de comunicaciones digital específico son de capital importancia. El ancho de banda del canal suele estar limitado para los canales de comunicaciones por dependencias gubernamentales. Por tanto, para aplicaciones con alta velocidad de transferencia de datos no suelen considerarse esquemas de señalización, tales como la FSK, en los que se utiliza un ancho de banda de modo ineficaz. El ancho de banda también determina el nivel de la potencia del ruido en el filtro de predetección del receptor, B_f .

Probabilidad de error. El diseñador también debe decidir qué proporción de error es capaz de tolerar la aplicación. Las ecuaciones de las tablas 48-20 y 48-21 pueden utilizarse a fin de calcular la probabilidad de error en los bits para los esquemas de señalización enumerados. Estos cálculos requieren el conocimiento de la función error complementario $Q(x)$ y de la métrica E_b/N_0 . $Q(x)$ puede calcularse así:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-u^2/2) du$$

Existen abundantes tablas en las que se enumera $Q(x)$ para un amplio intervalo del argumento x . Cuando se utilicen estas tablas para obtener valores de la función error complementario $Q(x)$, deben compensarse las diferentes definiciones empleadas.

Razón de señal sobre ruido. Como ya se mencionó P_E , la probabilidad de errores en los bits, es la métrica de mayor uso cuando se comparan diferentes esquemas de comunicaciones digitales. La métrica E_b/N_0 era la variable independiente en las ecuaciones para P_E . La SNR_i de predetección está directamente relacionada con esta cantidad. La SNR_s es una excelente función básica para evaluar el rendimiento del sistema. En la tabla se muestra esta cantidad para los diversos esquemas de señalización a una P_E fija igual a 10^{-4} . Utilizando la SNR_i como métrica ($P_E = 10^{-4}$), se observa a partir de los valores de la tabla y de la figura 48-143 que la PSK coherente es la que requiere la menor cantidad de potencia de los esquemas de modulación que se enumeran. En orden creciente de potencia requerida siguen la DPSK, la FSK coherente, la ASK incoherente, la ASK coherente y la FSK incoherente. De estos esquemas, los de mayor uso son la PSK, la DPSK y la FSK incoherente. A partir de la tabla se observa que la DPSK requiere aproximadamente 1 dB más de potencia que la PSK, y que la FSK incoherente requiere entre 6 y 7 dB más de potencia que la PSK.

Complejidad del equipo. Las diferencias más importantes con respecto a la complejidad del equipo para los esquemas de señalización digital en cuestión se presentan en el proceso de desmodulación. De las técnicas de modulación más frecuentemente utilizadas, la de

mayor complejidad es la del receptor PSK coherente. En orden decreciente de complejidad sigue la DPSK y por último la FSK incoherente.

Soluciones de compromiso en las aplicaciones. Es evidente que ninguno de los esquemas de señalización digital mencionados es el mejor para todas las aplicaciones. En las comunicaciones por satélite, la consideración más importante suele ser la potencia. La PSK coherente es la elección óptima para el sistema limitado en potencia, a pesar de la mayor complejidad del equipo. Por otra parte, cuando la potencia y velocidades de transferencia de datos no son las consideraciones de mayor importancia, a menudo una elección adecuada es la FSK incoherente.

48.7.4 Elección entre técnicas de modulación digital y analógica

Una respuesta a la pregunta de si se debe implementar un sistema de comunicaciones utilizando técnicas analógicas o bien digitales, depende de las aplicaciones. En las secciones anteriores se bosquejó una comparación de las técnicas de modulación analógica y digital. Aquí se analiza la alternativa entre ellas.

Se demostró que la salida de la fuente de información por lo general determina el tipo de modulación que es más fácil de implementar. Sin embargo, en algunos casos es deseable transmitir información analógica a través de un canal digital. En este caso es necesario convertir con exactitud la información analógica a información digital. El teorema del muestreo de Nyquist especifica que el número de muestras necesario para la reconstrucción exacta de la señal analógica es de dos muestras por ciclo a la mayor componente de frecuencia contenida en el proceso analógico limitado en banda. Luego se codifican estos valores muestrales en algún formato digital específico (modulación por codificación de pulsos, por ejemplo). El diseñador puede entonces aprovechar todas las técnicas disponibles de procesamiento de señales. Esta transmisión digital de información analógica incrementa la complejidad del equipo, además del ancho de banda de transmisión requerido.

Una aplicación tal, en la que la técnica podría justificar el costo en complejidad y ancho de banda, podría ser en un enlace espacial en el que se utilizaran repetidoras o transpondedores. En esta aplicación la información se desmodularía, se corregirían los errores y después aquélla volvería a modularse eliminando así el ruido de entrada de la repetidora. La TV es un ejemplo de señal analógica que a menudo es impráctico transmitir por medios digitales, debido al ancho de banda y a las velocidades del reloj requeridas.

48.7.5 Consideraciones prácticas

Sincronización

La sincronización es un concepto importante en los sistemas de comunicaciones analógicas y digitales. La complejidad del sistema a menudo determina el grado de sincronización requerido.

Tabla 48-21. Comparación de técnicas de modulación digital

Esquema	Símbolos, $(S_1(t), S_2(t))$	Ancho de banda de la velocidad de transferencia de los bits, W (Hz)	Probabilidad de error, P_E	S/N para $P_E = 10^{-4}$ (dB)	Complejidad del equipo	Utilización popular
Manipulación por desplazamiento de amplitud (ASK), coherente	$S_1(t) = A \cos \omega_c t$, $S_2(t) = \theta$	$2 W$	$Q(\sqrt{E_b/2N_0})^b$	11.4	Moderada	Se utiliza rara vez
ASK, incoherente	Los mismos que para la ASK coherente	$2 W$	$1/2 \exp(-E_b/2N_0)$	12.3	Secundaria	Requiere elevada potencia pico y es sensible a variaciones del canal
Manipulación por desplazamiento de frecuencia (FSK), coherente	$S_1(t) = A \cos$ $(\omega_c - \omega_d)t$, $S_2(t) = -A \cos$ $(\omega_c + \omega_d)t$	$> 2 W$	$Q(\sqrt{E_b/2N_0})$	10.6	Importante	Se utiliza poco
FSK, incoherente	Los mismos que para la ASK coherente	$> 2 W$	$1/2 \exp(-E_b/2N_0)$	12.3	Secundaria	Se utiliza para la transmisión de datos a baja velocidad
Manipulación por desplazamiento de fase (PSK), coherente	$S_1(t) = A \cos \omega_c t$, $S_2(t) = -A \cos \omega_c t$	$2 W$	$Q(\sqrt{2E_b/N_0})$	8.4	Importante	El mejor funcionamiento global para transmisión cuando se utiliza para datos a alta velocidad
PSK diferencial (DPSK)	Los mismos que para la PSK coherente con codificación diferencial	$2 W$	$1/2 \exp(-E_b/N_0)$	9.3	Moderada	Transmisión de datos a mediana velocidad

^a A = amplitud de la portadora; ω_c = frecuencia de la portadora, rad/s.

^b E_b = energía por bit; J ; N_0 = densidad espectral del ruido, W/Hz; $Q(x)$ función error complementario con argumento x .

Tanto los esquemas de desmodulación coherente analógicos como los digitales requieren al menos un nivel de sincronización. La sincronización de la portadora de los sistemas de comunicaciones digitales y analógicos requiere un conocimiento preciso de la referencia de fase de la portadora. La referencia de la portadora puede obtenerse de varias formas. Si el espectro de la señal de modulación contiene una componente portadora, entonces la coherencia de la portadora se obtiene utilizando un ciclo simple sincronizado ("enclavado") en fase. También es posible transmitir con la información un tono piloto relacionado de modo coherente con la portadora y recuperarlo para ser utilizado como referencia portadora en el receptor. Si el espectro de la señal de modulación no contiene una componente portadora, es posible utilizar técnicas opcionales, tales como un ciclo cuadrador o un ciclo de Costas, a fin de recuperar la portadora para desmodulación coherente. Para esquemas más complicados se requieren otros niveles de sincronización, como: sincronización de los bits, sincronización de palabras, sincronización de códigos.

Umbral

Una característica importante de algunos sistemas de comunicaciones es el efecto de umbral. La FM incoherente es un buen ejemplo de un sistema práctico que tiene un umbral muy definido, por debajo del cual el funcionamiento se degrada con rapidez (fig. 48-142). También es un excelente ejemplo de sistema útil para un amplio intervalo de requisitos en tanto se cuente con suficiente potencia para mantener el funcionamiento por encima del umbral. Si no se dispone de esa potencia, deben utilizarse algunas de las técnicas coherentes más complicadas y costosas, o bien debe incrementarse la complejidad del sistema FM para ampliar el umbral.

48.7.6 Resumen

En esta sección se ha presentado un breve repaso de los principios del análisis comparativo de las técnicas de comunicaciones. Estas técnicas analíticas son de utilidad en el diseño conceptual y en las etapas de planificación de un sistema de comunicaciones.

LIBROS Y REVISTAS

- W. R. Bennett, "Methods of Solving Noise Problems", *IRE* 44:609-638.
- W. R. Bennett, "Envelope Detection of a Unit-Index Amplitude Modulated Carrier Accompanied by Noise", *IEEE Trans. Inform. Theory* IT-20: 723-728.
- E. R. Berlekamp, *Algebraic Coding Theory*, McGraw-Hill, Nueva York.
- H. S. Black, *Modulation Theory*, Van Nostrand, Nueva York.
- A. B. Carlson, *Communication Systems*, McGraw-Hill, Nueva York.
- K. Feher, *Digital Communications, Microwave Applications*, Prentice-Hall, Englewood Cliffs, NJ.

- W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, Wiley, Nueva York.
- R. M. Gagliardi, *Introduction to Communication Engineering*, Wiley, Nueva York.
- F. M. Gardner, *Phaselock Techniques*, Wiley, Nueva York.
- W. D. Gregg, *Analog and Digital Communications*, Wiley, Nueva York.
- Y. W. Lee, *Statistical Theory of Communications*, Wiley, Nueva York.
- W. C. Lindsey y M. K. Simon, *Telecommunication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ.
- R. W. Lucky, "A Survey of the Communication Theory Literature: 1968-1973", *IEE Trans. Inform. Theory* IT-19:725-739.
- D. Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill, Nueva York.
- P. F. Panter, *Modulation, Noise, and Spectral Analysis*, McGraw-Hill, Nueva York.
- P. F. Panter, *Communication Systems Design: Line-of-Sight and Troposcatter Systems*, McGraw-Hill, Nueva York.
- A. Papoulis, *The Fourier Integral and Its Application*, McGraw-Hill, Nueva York.
- A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Nueva York.
- S. O. Rice, "Mathematical Analysis of Random Noise", *Bell Sys. Tech. J.* 24:46-156.
- D. J. Sakrison, *Communication Theory: Transmission of Waveform and Digital Information*, Wiley, Nueva York.
- M. Schwartz, *Information Transmission, Modulation and Noise*, 2.^a ed., McGraw-Hill, Nueva York.
- M. Schwartz, W. R. Bennett y S. Stein, *Communication Systems and Techniques*, McGraw-Hill, Nueva York.
- R. J. Schwartz y B. Friedland, *Linear Systems*, McGraw-Hill, Nueva York.
- K. S. Shanmugam, *Digital and Analog Communication Systems*, Wiley, Nueva York.
- C. E. Shannon, "A Mathematical Theory of Communications", *Bell Sys. Tech. J.* 27:379-423, 623-656.
- B. Sklar, "A Structured Overview of Digital Communications", *IEEE Commun. Mag.*
- M. I. Skolnik, ed., *Radar Handbook*, McGraw-Hill, Nueva York.
- J. J. Spilker, *Digital Communications by Satellite*, Prentice-Hall, Englewood Cliffs, NJ.
- H. Stark y F. B. Tuteur, *Modern Electrical Communications Theory and Systems*, Prentice-Hall, Englewood Cliffs, NJ.
- H. Taub y D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, Nueva York.
- N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, Wiley, Nueva York.
- Wozencraft y Jacobs, *Principles of Communication Engineering*, Wiley, Nueva York.
- R. Ziemer y W. H. Tranter, *Principles of Communications, Systems, Modulations, and Noise*, Houghton Mifflin, Boston.