

818

R621.381  
R452e  
V.7, ej.1

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

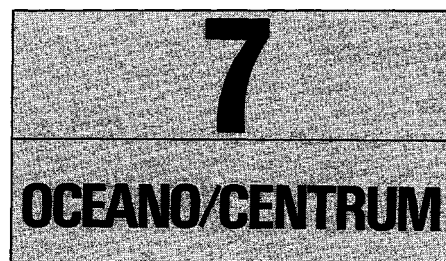
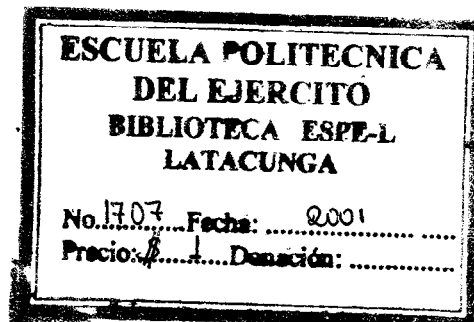
**Charles Belove**

**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.<sup>a</sup> 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

**Milton Rosenstein**

New York Institute of Technology  
Old Westbury, Nueva York

**Edward Nelson**

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

**Juan Carlos Vega Fagoaga**

Ingeniería en Sistemas

**Hugo Villagómez Velázquez**

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

**José Rafael Blengio Pinto**

Médico Cirujano  
Universidad Nacional Autónoma  
de México

**M.<sup>a</sup> 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 7: 84-7841-023-6

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

## DÉCIMA PARTE (CONTINUACIÓN) COMUNICACIONES

### 49. Sistemas de comunicación, 1621

Ray W. Nettleton, Jefferson F. Lindsey  
III, Surya V. Varanasi, Daniel F.  
DiFonzo

Sistemas terrestres, 1622  
Sistemas de radiodifusión, 1640  
Redes de comunicaciones, 1661  
Sistemas de comunicaciones  
por satélite, 1672

## UNDÉCIMA PARTE SISTEMAS DE TELEMETRÍA, NAVEGACIÓN Y ATERRIZAJE

### 50. Principios de navegación, 1711

Emil R. Schiesser

Definiciones de funciones, 1712  
Propagación del estado, 1712  
Determinación del estado, 1713  
Observación y medición, 1713  
Sistemas de navegación, 1714  
Resumen, 1715

### 51. Radar, 1717

Frank R. Castella

El problema de la navegación, 1718  
Principios del radar, 1718  
Tipos de radar, 1719

Ecuaciones de funcionamiento  
del radar, 1720

Resolución del radar ideal, 1722

Limitaciones del sistema en cuanto  
a exactitud de la medición, 1723

Efectos de la trayectoria  
de propagación, 1725

Radar Doppler y mediciones  
de velocidad, 1727

Técnicas avanzadas (ondas  
milimétricas), 1729

### 52. Auxiliares de la navegación, 1731

Frederick B. Pogust, Stephen C. Martin  
y Thomas J. Cutler

Normas y convenios internacionales  
de navegación aérea, 1732

Auxiliares de la navegación aérea, 1737

Auxiliares de la navegación marítima, 1764

### 53. Sistemas de aterrizaje de aeronaves, 1777

Stephen C. Martin

Radioayuda para el aterrizaje  
de aeronaves, 1778

Sistema de aterrizaje por instrumentos  
(ILS), 1778

Radar de precisión para el aterrizaje  
(PAR), 1784

Sistema de aterrizaje por microondas  
(MLS), 1784

Sistemas de aterrizaje militares para fines  
especiales, 1789



**54. Navegación por satélite, 1793**

John H. Painter

Principios de la radionavegación por satélite, 1794

Descripción de la órbita de un satélite (ecuación de Kepler), 1795

Coordenadas del usuario y geodesia, 1797

Adquisición de datos de navegación, 1798

Efectos ionosféricos, 1799

TRANSIT, sistema Doppler de baja dinámica, 1802

GPS NAVSTAR, sistema telemétrico de alta dinámica, 1806

**DECIMOSEGUNDA PARTE  
COMPUTADORAS**

**55. Sistemas numéricos y representación de datos, 1831**

Alan B. Marcovitz

Números binarios, octales y hexadecimales, 1832

Suma binaria de números positivos, 1837

Números con signo, 1837

Aritmética binaria, 1839

Decimal codificado en binario (BCD), 1841

Aritmética de precisión múltiple, 1844

Aritmética de punto flotante, 1844

Detección y corrección de errores, 1847

Representaciones de datos alfanuméricos, 1849

**56. Organización y arquitectura de la computadora, 1851**

Edward J. Lancevich

Organización general (diagrama de bloques), 1852

Registros de la CPU, 1853

Tipos de instrucciones, 1854

Formatos de instrucciones y modos de direccionamiento, 1859

Modos de operación, 1867

Diseño de la unidad de control, 1872

Clasificación de los sistemas de computación, 1876

Comparación de computadoras, 1880

**57. Estructuras de datos, 1881**

James A. M. McHugh

Introducción, 1882

Listas lineales enlazadas, 1882

Líneas de espera y pilas, 1885

Árboles, 1888

Tablas de asignación, 1892

## CAPÍTULO 49

# Sistemas de comunicación

### **Ray W. Nettleton**

AMECOM Division of Litton Industries, Inc.  
College Park, Maryland

### **Jefferson F. Lindsey III**

Southern Illinois University at Carbondale  
Carbondale, Illinois

### **Surya V. Varanasi**

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

### **Daniel F. DiFonzo**

COMSAT Laboratories  
Clarksburg, Maryland

#### **49.1 Sistemas terrestres**

- 49.1.1 Introducción
- 49.1.2 Propagación de la señal
- 49.1.3 Radioenlaces de punto a punto
- 49.1.4 Radioenlaces móviles

#### **49.2 Sistemas de radiodifusión**

- 49.2.1 Radiodifusión estándar (modulación de amplitud)
- 49.2.2 Modulación de frecuencia (FM)
- 49.2.3 Emisiones televisivas

#### **49.3 Redes de comunicaciones**

- 49.3.1 Introducción
- 49.3.2 Consideraciones para el diseño de redes
- 49.3.3 Terminales de datos

- 49.3.4 Configuraciones de redes y conmutación de datos

- 49.3.5 Optimización de circuitos y transferencia efectiva de datos
- 49.3.6 Consideraciones operacionales
- 49.3.7 Protocolos

#### **49.4 Sistemas de comunicaciones por satélite**

- 49.4.1 Introducción y repaso
- 49.4.2 Configuraciones geométricas de órbitas y lanzamientos
- 49.4.3 Parámetros de enlace de los sistemas de comunicación por satélite
- 49.4.4 Sistemas de satélite representativos
- 49.4.5 Subsistemas
- 49.4.6 Modulación y acceso
- 49.4.7 Temas diversos

## 49.1 SISTEMAS TERRESTRES

Ray W. Nettleton

### 49.1.1 Introducción

Esta sección se dedica a la transmisión de información entre dos puntos, ambos en la Tierra y sin trayectoria de conducción entre ellos. Aquí no se aborda el empleo de mecanismos extraterrestres, tales como satélites; para ello debe consultarse la sección 49.4, que trata de las comunicaciones por satélite.

### 49.1.2 Propagación de la señal

La propagación de una señal de radio entre dos puntos terrestres se basa en la interacción de las ondas de radio con la atmósfera de la Tierra, la superficie de ésta y cualesquiera obstrucciones y desuniformidades que pudieran presentarse.

Es posible identificar cinco mecanismos principales de propagación:

1. En línea recta.
2. De ondas terrestres.
3. Ionosférica.
4. Por dispersión troposférica.
5. Por dispersión local<sup>1,2</sup> (fig. 49-1).

Cada uno de estos mecanismos crea un canal con sus propias características que varían con la frecuencia, distancia de enlace, hora del día y año, así como con las condiciones solares y atmosféricas. A menudo actúan juntos varios mecanismos, pero por lo general la intensidad de la señal debida a uno de ellos domina a los demás, y es posible modelar el canal como si un solo mecanismo estuviera en operación. El mecanismo dominante de transmisión puede cambiar de un lugar a otro (especialmente si la longitud de la trayectoria cambia de manera significativa), y algunas veces de una ocasión a otra en el mismo lugar.

El diseño de un enlace de comunicación depende de muchos factores, algunos bajo el control del diseñador (tales como la relación costo/rendimiento y el método de modulación) y otros fuera de su control. Para el caso de enlaces terrestres no guiados, el factor de diseño no controlado de mayor importancia es la naturaleza del canal por el que debe efectuarse la transmisión. Por tanto, en esta sección se aborda la naturaleza de cada canal y se describen algunas aplicaciones prácticas para enlaces fijos de punto a punto y para radioenlaces móviles.

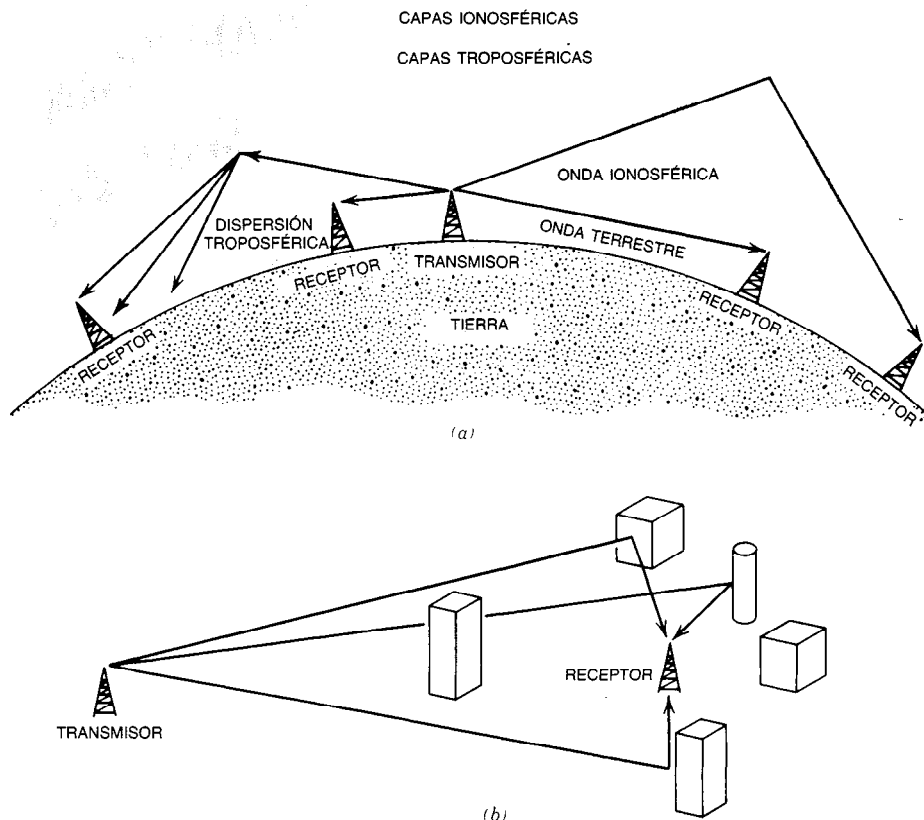


Fig. 49.1. Principales mecanismos de propagación: a) propagación atmosférica y terrestre; b) propagación por dispersión local.

### Propagación en línea recta

Un canal de línea recta o de trayectoria óptica directa (LOS, de *line-of-sight*) puede establecerse siempre que una trayectoria en línea recta entre el transmisor y el receptor se encuentre libre de obstrucciones importantes. El término "importantes" aquí significa cualquier obstrucción cuyas dimensiones sean grandes comparadas con una longitud de onda. Para ondas de radio esto suele significar que es posible ignorar obstrucciones pequeñas pero visibles (como las hojas de un árbol), de forma que podría existir una trayectoria LOS incluso si un observador colocado en el transmisor no pudiese ver realmente el receptor. Además, la presencia de obstrucciones importantes cerca de la trayectoria de transmisión puede ocasionar la distorsión de la señal en múltiples trayectorias, por lo que en la mayor parte de los enlaces LOS se emplean patrones de antenas estrechas enfocadas mutuamente a fin de evitar problemas de trayectoria múltiple. Esto permite también el uso múltiple de la misma banda de frecuencia en sistemas independientes pero cercanos. A fin de poder emplear antenas de haz estrecho de tamaño físico razonable se requieren pequeñas longitudes de onda. Así, en la mayor parte de los enlaces LOS se utilizan frecuencias de microondas, milimétricas y ópticas, aunque teóricamente tales enlaces pueden establecerse en cualquier frecuencia.

Sin embargo, a pesar del diseño y la ubicación cuidadosos del equipo de enlace LOS, la Tierra y la atmósfera a menudo propician la recepción de trayectoria múltiple.

En la figura 49-2 se muestran dos posibilidades. La primera se presenta en condiciones por lo demás perfectas; una parte del frente de onda transmitido es reflejada por la superficie de la Tierra, por otras obstrucciones fijas o por ambas, e incide en la antena del receptor. Dado que las ondas dirigida y reflejada han viajado distancias diferentes, suelen no estar en fase; si son de la misma magnitud, incluso podrían cancelarse entre sí y producir un campo neto recibido igual a cero. Por consiguiente, existe una altura óptima de antena receptora para cualquier enlace, que corresponde a la primera zona de Fresnel que se abre desde el transmisor, y que depende de la geometría transmisor-tierra-receptor y del desfase en la reflexión (es decir, la más alta no necesariamente es la mejor; obsérvese que las antenas de microondas a menudo están montadas en alguna parte intermedia de una torre o de un edificio alto, en vez de estarlo en la parte superior).

El primer mecanismo de trayectorias múltiples es fácil de explicar, ya que es invariante con el tiempo. El segundo es más difícil: la refracción de parte de la onda transmitida en regiones de la atmósfera cuyos índices de refracción son diferentes. Esto ocurre debido a la presencia de capas térmicas y turbulencias, de forma que el fenómeno cambia con la hora del día y con el clima. Las señales recibidas pueden estar en fase o parcial o totalmente canceladas. El marco temporal del fenómeno varía bastante; en los enlaces de microondas, para completar un ciclo se requieren minutos u horas, y el efecto es un "desvanecimiento". En enlaces ópticos esto puede ocurrir muchas veces por segundo, y el resultado se denomina centelleo.

La forma más común de combatir estos desvanecimientos consiste en proporcionar diversidad de espacio; es decir, múltiples antenas receptoras separadas entre sí algunas longitudes de onda, de forma que los desvanecimientos encontrados por cada una tiendan a no correlacionarse estadísticamente. En el caso de las microondas esto significa dos antenas montadas a diferentes alturas de la torre. Para frecuencias ópticas, basta incrementar el tamaño de la superficie del fotodetector, de forma que la suma de éste corresponda a la suma de varios modos ópticos recibidos.

Un segundo problema fundamental para el diseñador de enlace LOS es la presencia de precipitación en la trayectoria de propagación. Lluvia, nieve y niebla producen atenuación de las ondas de radio. El efecto es mayor para la precipitación más densa, partículas más grandes de precipitación (p. ej., la nieve es peor que la niebla, y el granizo suele ser el más perjudicial), y las frecuencias superiores. A frecuencias ópticas incluso el polvo y otras partículas atmosféricas (aerosoles) representan un problema. En la sección correspondiente a enlaces de microondas y ópticos se presentan curvas de pérdidas específicas.

### Propagación de ondas terrestres

Debido a que la tierra es un conductor imperfecto, las ondas de radio penetran algo en su superficie; su intensidad de campo disminuye rápidamente con la profundidad, y se propagan más lentamente que en el aire. Esto origina un "arrastre" o efecto de onda superficial, que provoca que, justo por encima de la superficie, la onda siga la curvatura de la Tierra. Este proceso es inherentemente disipativo, y es de utilidad sólo para frecuencias relativamente bajas y para comunicación transhorizonte a distancias relativamente cortas, del

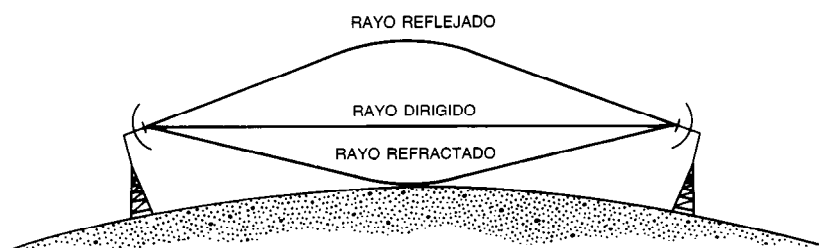


Fig. 49-2. Propagación por trayectorias múltiples en enlaces en línea recta (visual).

orden de 200 km para baja frecuencia media (MF, de *medium frequency*) y de 2 000 km para baja frecuencia (LF, de *low frequency*) y muy baja frecuencia (VLF, de *very low frequency*).

La propagación de ondas terrestres suele presentarse con otras formas de propagación, y provoca patrones de interferencia que motivan fenómenos de desvanecimiento. Es útil durante el día, cuando la capa D fuertemente absorbente de la ionosfera hace muy débil la reflexión en las capas superiores de las frecuencias baja y mediana. También se utiliza para comunicaciones marinas y submarinas a LF, VLF y frecuencia extremadamente baja (ELF, de *extremely low frequency*) (<300 kHz)<sup>3</sup>.

#### Propagación ionosférica

La ionosfera es una región de la atmósfera superior que contiene capas de gases atmosféricos cargados eléctricamente. La ionización de los gases se debe a la acción de la radiación solar (tanto la radiación electromagnética como el denominado viento solar, que es una corriente de partículas ionizadas lanzadas a gran velocidad desde el Sol). De este modo, la composición de la ionosfera depende en gran medida de la actividad solar, particularmente de las manchas solares y de la hora del día (variación diurna), de la estación del año (variación estacional) y de la latitud geográfica.

En la figura 49-3 se muestra la variación diurna en la ubicación de las principales capas de la ionosfera. Durante el día hay cuatro capas diferentes: D, E, F<sub>1</sub> y F<sub>2</sub>. Durante la noche desaparece la capa D y se presenta la capa F. Se reduce la densidad de la capa E, lo cual permite mayor penetración en la capa F. Además de

las capas más predecibles, pequeñas capas parecidas a nubes de densidad iónica incrementada vagan alrededor de la capa E, permitiendo una breve comunicación mejorada, por lo general en la banda superior HF. Estas nubes se denominan E esporádicas.

La ionosfera también es afectada por el mismo campo magnético de la Tierra. La presencia de este campo da a la ionosfera la propiedad de birrefringencia, por lo que separa en dos componentes una onda de radio incidente, una polarizada a lo largo de las líneas de fuerza magnéticas y otra normal a ellas. Las ondas resultantes reflejadas se denominan ondas ordinarias y extraordinarias, y sus características pueden ser bastante diferentes. La actividad de la corona solar también provoca tormentas magnéticas en la ionosfera, especialmente en los polos de la Tierra, lo que a menudo se manifiesta en las célebres y espectaculares auroras boreales (luzes del Norte). Tales tormentas hacen estragos en los canales de radio ionosféricos, provocando con frecuencia bloqueo total en las radiocomunicaciones.

Las partículas ionizadas vibran físicamente en respuesta a la presencia de un campo de radiofrecuencia, y la absorción de la energía resultante de las colisiones de los iones produce atenuación en la onda de radio. Esta absorción alcanza su máxima anchura a una frecuencia resonante específica para cada especie molecular (la frecuencia giromagnética)<sup>4</sup>, por lo general a frecuencias por debajo de 1 kHz para los gases atmosféricos (banda ELF). Sin embargo, el ion de mayor importancia para la propagación HF es el electrón libre, cuya resonancia giromagnética es de aproximadamente 1.4 MHz. Así, la absorción es un máximo alrededor del

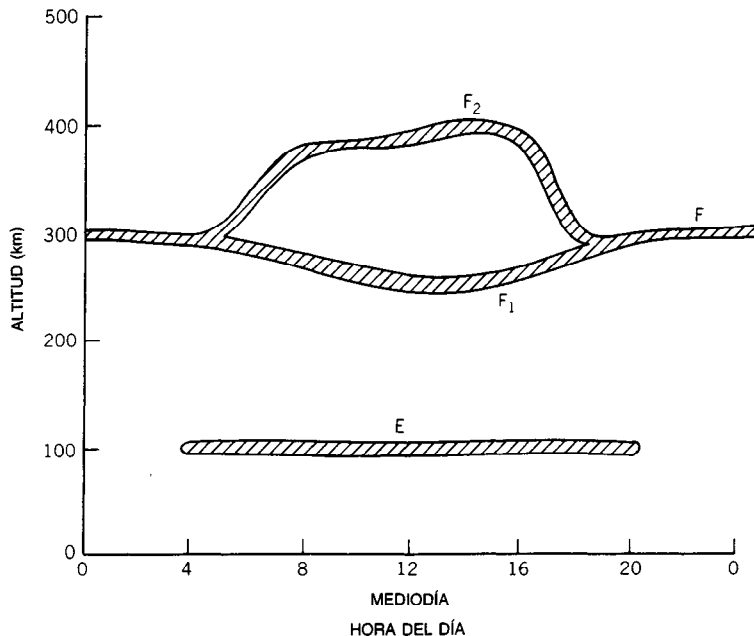


Fig. 49-3. Variaciones diurnas en las capas ionosféricas. Adaptado de la Circular 462 de National Bureau of Standards, "Ionospheric Radio Propagation".

centro de la banda MF (300 kHz-3 MHz). La absorción también depende de la densidad de los iones y es un máximo en la densa capa D. Cuando esta capa desaparece en la noche, ocurre algo de propagación en la banda MF. Sin embargo, la atenuación es gradualmente menor a frecuencias por encima y por debajo de la región de resonancia, y los efectos de reflexión más útiles se dan en las frecuencias LF y VLF (<300 kHz) y en el intervalo de la alta frecuencia (HF, de *high frequency*) (3-30 MHz y algunas veces más allá) por debajo de la máxima frecuencia utilizable (MUF, de *maximum usable frequency*) en la actualidad para la capa que se esté empleando.

La "frecuencia crítica" de una capa ionosférica es la mayor frecuencia que será reflejada de dicha capa a un ángulo normal (vertical) de incidencia. La frecuencia crítica depende en gran medida de las variaciones diurna, estacional y de los ciclos de las manchas solares. En la figura 49-4 se muestran las variaciones diurnas y el efecto del ciclo de las manchas solares en la frecuencia crítica.

La propagación por reflexión ocurre por encima de la frecuencia crítica a ángulos oblicuos de incidencia. La máxima frecuencia utilizable (MUF) a un ángulo de incidencia  $\phi$  está dada por:

$$MUF = f_c \sec \phi$$

donde  $f_c$  es la frecuencia crítica.

Dado que la absorción empieza a aumentar a medida que se reduce la frecuencia por debajo de la MUF, la frecuencia de operación suele elegirse justo por debajo de la MUF, dejando un margen adecuado para errores debidos a condiciones cambiantes y predicción ionosférica inexacta. La National Bureau of Standards del U.S. Department of Commerce<sup>5</sup> y otras agencias en Estados Unidos y diversos países publican

tablas de datos para la frecuencia óptima de operación en función de la localización geográfica, la hora del día y la fecha. Por ejemplo, los informes más inmediatos de las condiciones de propagación en el momento se emiten cada hora, exactamente 14 minutos después de pasada la hora (hora media del meridiano de Greenwich), por la estación de radio WWV, Fort Collins, Colorado, a 2.5, 5, 10, 15, 20 y 25 MHz. La misma información puede obtenerse llamando a dicha estación al (303)499-7111 (Estados Unidos).

Con estos argumentos también se concluye que en la comunicación entre la Tierra y el espacio extraterrestre (incluyendo las comunicaciones tierra-tierra vía satélite) deben utilizarse frecuencias por *arriba* de la MUF al ángulo agudo de interés, ya que tales señales deben pasar todas las capas de la ionosfera.

Una consecuencia directa de la práctica común de elegir una frecuencia de operación entre la frecuencia crítica y la MUF es que el transmisor se encuentra rodeado por una "zona de silencio" aproximadamente anular en la que la recepción es difícil o casi imposible. Al alejarse del transmisor, la zona empieza donde la intensidad de campo de las ondas terrestres ha caído por debajo del umbral de recepción, y termina donde la primera onda ionosférica haya regresado a la Tierra. En la figura 49-5 se ilustra la noción de "distancia de salto", que es el ancho de la zona medido radialmente.

Una consecuencia secundaria del efecto de las ondas de radio en la ionosfera es que un campo de radio suficientemente intenso es capaz de provocar una perturbación medible de la ionosfera en sí. El fenómeno más celebrado en este sentido es el denominado efecto Luxemburgo, en el que la señal de una estación alejada se observa como si hubiera sido modulada por el mensaje de una poderosa estación cercana de radio AM.

El desvanecimiento es un efecto especialmente problemático en muchos casos de propagación ionosférica.

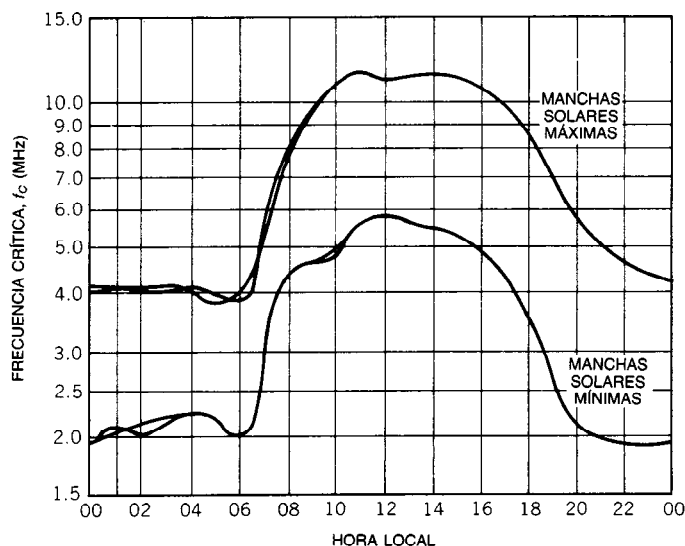


Fig. 49-4. Variación diurna de la frecuencia crítica. Adaptado de la Circular 462 de National Bureau of Standards, "Ionospheric Radio Propagation".

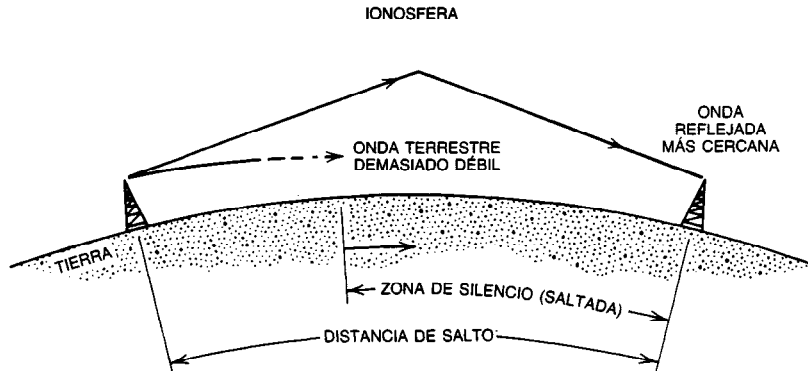


Fig. 49-5. Fenómeno de "salto" en la propagación ionosférica.

rica. Es producido por varios mecanismos. El primero y más común de los problemas es el carácter no estacionario de las capas ionosféricas mismas. Como resultado de fenómenos celestes y meteorológicos ocurren constantemente movimientos físicos, tanto horizontales como verticales, y cambios en densidad. Por tanto, los cambios a largo plazo en las características de reflexión son la regla más que la excepción. Este tipo de desvanecimiento es particularmente observable al amanecer y en el crepúsculo, que es cuando la estructura ionosférica cambia radicalmente y la frontera de una zona de silencio puede oscilar de un lado a otro de un receptor. Este desvanecimiento se denomina desvanecimiento MUF.

Sin embargo, los cambios a corto plazo son más observables y ocurren debido a la recepción de trayectoria múltiple. La onda recibida puede ser la suma de varias ondas incidentes, que llegan en ángulos ligeramente distintos y con diversas longitudes de trayectoria. El resultado es un desvanecimiento tipo interferencia con estadísticas de Rayleigh o de Rice y una intensidad de campo que varía en periodos relativamente cortos. En la tabla 49-1 se muestran las características de desvanecimiento para diferentes frecuencias y longitudes de onda.

#### Dispersión troposférica

La región de la atmósfera donde vive el ser humano se denomina troposfera. Se caracteriza por una temperatura que en general disminuye con la altitud, y por capas alternadas de flujo de aire turbulento y laminar a temperaturas significativamente diferentes. Los índices de refracción de estas capas son tan diferentes que se experimenta algo de dispersión de regreso a la Tierra en las bandas de muy alta frecuencia (VHF, de *very-high-frequency*) y ultraalta frecuencia (UHF, de *ultra-high-frequency*), hasta las frecuencias en las que la absorción de gases y vapor atmosféricos comienza a provocar seria atenuación (resonancia dipolar molecular). Este fenómeno, conocido como dispersión troposférica o tropodispersión, tiene un intervalo utilizable de alrededor de 30 MHz a 10 GHz, y casualmente comienza donde termina la propagación ionosférica.

Comparada con la propagación ionosférica, la tro-

podispersión es mucho menos propensa a cambios en las condiciones solares o estacionales, aunque varía muy notablemente con la ubicación geográfica, y es de mayor utilidad en las zonas templadas que en los trópicos. Dado que la dispersión ocurre a altitudes reducidas, el intervalo está más limitado (del orden de los cientos de kilómetros). La dispersión troposférica también adolece de los fenómenos de trayectoria múltiple y desvanecimiento. El número de trayectorias discernibles para la onda recibida suele ser muy elevado, lo que da por resultado la bien conocida función densidad de probabilidad de Rayleigh para la intensidad de campo recibida:

$$p(E) = (E/P) \exp(-E^2/2P)$$

donde  $E$  es la magnitud del campo eléctrico recibido y  $P$  es la potencia media recibida en el corto plazo, que varía a su vez en grandes periodos de tiempo y tiene una distribución logarítmica normal:

$$p(P_d) = (2\pi S)^{-1/2} \exp[-(P_d - Q)^2/2S]$$

donde  $P_d$  = media a corto plazo, dB

$Q$  = media global, dB

$S$  = cuadrado de la desviación estándar de la distribución, típicamente alrededor de 2 a 5 db

El desvanecimiento a corto plazo debido a la trayectoria múltiple y a la turbulencia se denomina desvanecimiento rápido, y el de la media a corto plazo (debido a variaciones diurnas y meteorológicas en la estructura de capas de la troposfera), desvanecimiento lento.

A fin de reducir la probabilidad de que la señal recibida caiga por debajo del umbral de recepción durante un profundo desvanecimiento, se incluye un "margen de desvanecimiento" en el cálculo de la potencia transmitida requerida. Si es posible lograr un funcionamiento satisfactorio *sin* desvanecimiento lento y con una potencia mínima transmitida de  $T$  dBW, entonces la probabilidad de rendimiento insatisfactorio *con* desvanecimiento lento cuando la potencia transmitida sea  $T + mS$  dBW es  $\frac{1}{2} \operatorname{erfc}(m/\sqrt{2})$ , donde  $\operatorname{erfc}(\cdot)$  es la

Tabla 49-1. Velocidades de desvanecimiento en función de la frecuencia (orden de magnitud)

$f$	Distancia de transmisión (km)	Hora del día	Tipo de desvanecimiento	Profundidad del desvanecimiento (%)	Periodo
20 kHz	7000	De noche	Interferencia	5-25	30-60 min
20 kHz	0	De noche	Interferencia	50	10 min
20 kHz	Variable	De día	Interferencia	0	Sin desvanecimiento apreciable
100 kHz	0	De noche	Interferencia	50	1 a 5 min
1 MHz	1 000	De noche	Interferencia	85	0.1-1.0 min
1 MHz	1 000	De día	Interferencia	0	Sin desvanecimiento apreciable
10 MHz	0-10 000	Noche y día	Interferencia	100	1 a 10 s
10 MHz	3 000	Por la tarde	Fluctuación	?	0.1 a 0.01 s
100 MHz	Oblicua	Más común por la noche	Escintilación de las radioestrellas		10 a 100 s
100 MHz	Ecos lunares	Noche y día	Libración	Pequeña	1 a 100 s
			Polarización	Grande	30 min
20 MHz	Satélite	Noche y día	Escintilación	90	0.2 s
20 MHz	Satélite	Noche y día	Polarización	Grande	1 s

Fuente: K. Davies<sup>4</sup>.

función error complementario. Por ejemplo, para una desviación estándar de 3 dB, una probabilidad de rendimiento satisfactorio del 99.7% requerirá  $m = 3$  o un margen de desvanecimiento de 9 dB.

El desvanecimiento rápido también puede combatirse mediante el empleo de recepción múltiple (diversidad). La recepción múltiple con una separación de antenas del orden de 100 longitudes de onda suele ser suficiente. Debido a la naturaleza de trayectoria múltiple de la señal recibida, es improbable obtener la ganancia total de una gran antena, ya que diferentes porciones del frente de onda recibido estarán fuera de fase sobre una gran área de recepción (frente de onda incoherente). Por tanto, con aparatos electrónicos apropiados, es probable que dos o más pequeñas antenas se comporten cuando menos tan bien como una sola antena grande en los receptores de dispersión troposférica, y a menor costo.

#### Propagación por dispersión local

Este mecanismo de propagación no pertenece al grupo de los canales "clásicos" ya descritos, por lo que no se menciona en la mayor parte de los manuales de este tipo. Sus características no se han investigado por completo hasta hace muy poco, aunque su existencia es tan antigua como la radiocomunicación. Está tipificada por el canal de radioenlaces móviles urbanos<sup>2,6</sup>, pero otros ejemplos, tales como el canal acústico submarino<sup>7</sup>, tienen propiedades muy parecidas.

Cuando la trayectoria de enlace es relativamente corta, no es necesario basarse en efectos atmosféricos de gran altitud o en la propagación de ondas terrestres, incluso cuando no exista trayectoria en línea recta (visible) entre el transmisor y el receptor. Cuando el receptor se encuentra rodeado de obstrucciones, son

ellas mismas (edificios, postes para el cableado eléctrico, árboles, colinas) las que proporcionan una multitud de trayectorias para la señal de radio.

La señal recibida de este canal llega desde muchas direcciones y por varios mecanismos. En el caso de grandes obstrucciones (en comparación con una longitud de onda), las ondas pueden reflejarse, difractarse o ambas cosas. En el caso de pequeñas obstrucciones ocurrirá dispersión, y cada partícula actuará como un débil radiador omnidireccional. Afortunadamente, por lo general no es necesario considerar los mecanismos específicos que actúan en un canal particular de este tipo. El efecto neto es como si la antena receptora estuviese rodeada por un gran número de dispersores colocados aleatoriamente, actuando cada uno de ellos como una fuente de la misma señal pero con fase, retardo y amplitud aleatorios. El campo recibido resultante tiene amplitud con distribución de Rayleigh y fase distribuida uniformemente. Sus estadísticos son esencialmente independientes para muestras tomadas con separaciones de varias longitudes de onda (a la "distancia de coherencia" lateral o más allá), para muestras tomadas a dos frecuencias lo suficientemente separadas (espaciadas por, al menos, el "ancho de banda de coherencia" del canal), y siempre que ocurra movimiento, ya sea en el transmisor, el receptor o los dispersores, para muestras cuya separación temporal sea lo suficientemente grande (por lo menos la "coherencia temporal" del canal).

Ocasionalmente, una sola de las trayectorias producirá una intensidad de campo que domine a las demás en cuanto a magnitud. Esto puede ocurrir en localizaciones "abiertas", tales como en la proximidad de un parque o de un lago, donde una trayectoria LOS puede estar acompañada por muchas trayectorias dis-



persas; o bien cuando una gran superficie provoque una reflexión particularmente fuerte. Tales trayectorias dominantes se denominan trayectorias especulares y son semejantes al bien conocido fenómeno de radar denominado reflexión fluctuante. En este caso el campo resultante es la suma fasorial del campo especular y el campo disperso, y tiene una función densidad de probabilidad con distribución de Rice:

$$p(E) = (E/S) \exp [-(E^2 + P^2)/2S] I_0(EP/S)$$

donde  $I_0(\cdot)$  = función de Bessel de primera clase y orden cero

$E$  = magnitud del campo eléctrico recibido

$P$  = media a corto plazo de la amplitud especular recibida

$S$  = media a corto plazo de la potencia total recibida

La brevedad del tiempo de promediación depende del tiempo que continúe el canal exhibiendo estadísticos cuasiestacionarios. En el caso de un auto que se desplaza con gran velocidad en un canal de radio móvil, esto podría ser de sólo unos cuantos segundos. Se dice que tal canal experimenta desvanecimiento de Rice que, como era de esperar, es menos grave que el desvanecimiento de Rayleigh.

#### Otros fenómenos de propagación

Algunas veces, pueden ser responsables de la recepción de señales otros fenómenos de propagación diferentes a los cinco que se acaban de describir. Algunos de ellos son los siguientes.

**Propagación por estelas meteóricas.** Esto ocurre debido al bombardeo de la atmósfera por meteoritos, lo que provoca columnas cortas, pero densas, de partículas ionizadas en la ionosfera, especialmente en la capa E. Las señales de VHF y UHF se reflejan en tales columnas y es posible que sean recibidas hasta una distancia de 1 000 km. Tales canales, aunque de corta duración, pueden utilizarse para la comunicación práctica de los denominados mensajes en ráfaga o de bajo régimen de trabajo, especialmente durante periodos de tormentas de meteoros en primavera y otoño.

**Propagación guiada por conductos atmosféricos.** Esto ocurre cuando las capas atmosféricas están dispuestas de forma que producen una "guía de ondas" efectiva en el espacio entre ellas. Se cree que los conductos se forman entre las capas de la ionosfera (mejorando la propagación HF a larga distancia), dentro de la troposfera, y entre la Tierra misma y la parte inferior de la capa D (durante el día) o de la capa E (en la noche), de modo que se produce una cobertura casi global de las ondas VLF y ELF.

**Propagación por difracción.** Esto ocurre siempre que una onda pasa sobre el borde de una obstrucción. La teoría clásica de la difracción se basa en la "agudeza infinita" del borde, por lo que se requiere alguna corrección en caso de que el borde no sea lo suficientemente agudo. En un principio se efectuaron algunos

esfuerzos, en el desarrollo de las radiocomunicaciones, para explicar la propagación transhorizonte por medio de la teoría de la difracción, y en muchos manuales todavía aparecen extensos análisis acerca de ella. El hecho es que, aunque la difracción ocurre en la mayor parte de los casos de propagación, sólo en algunos muy especiales explica de forma predominante la intensidad de campo presente en un receptor. Probablemente la ocurrencia más común es en la difracción sobre la parte superior de una cordillera hasta el "área sombreada" situada detrás. Paradójicamente, la difracción es importante en la propagación LOS, donde la trayectoria del rayo está cerca de la superficie de la Tierra (consúltese el análisis de los enlaces LOS de microondas).

**Cinturones dipolares orbitales.** Quizá el mayor interés de éstos sea de carácter histórico. En 1963, el West Ford Experiment puso en órbita a 3 700 km de la Tierra aproximadamente un cinturón de pequeños dipolos de más o menos 1.8 cm de longitud, en un intento por obtener un medio reflector de microondas para comunicación a larga distancia. Experimentos de transmisión demostraron que la idea era marginalmente factible, pero las protestas de círculos científicos, especialmente entre los radioastrónomos, así como el éxito de satélites de comunicación activos, hicieron aparcarse el experimento. Se duda que llegue a repetirse algún experimento como éste.

#### 49.1.3 Radioenlaces de punto a punto

En las siguientes secciones se describen los problemas específicos, consideraciones de diseño y prácticas de radioenlaces en los cuales las terminales transmisora y receptora están ubicadas en puntos fijos de la Tierra. Si las secciones tienen acaso algo en común, es el hecho de que, por lo general, en el enlace no ocurre un efecto Doppler significativo; sin embargo, la generalidad de tales enlaces todavía presentan características que varían en el tiempo. El análisis está organizado en orden creciente de frecuencia. Dado que las bandas de frecuencia en la MF y por debajo de ella se utilizan para la radiodifusión y aplicaciones móviles, de navegación y de frecuencia estándar más que para las comunicaciones, se empezará en la banda HF.

##### Radioenlaces en HF

La banda HF (3 y 30 MHz) se emplea para enlaces de mediano y largo alcance en los que la cobertura del satélite o de la línea de tierra es inexistente o no es económica. Aplicaciones típicas son la telegrafía, la telefonía y transmisión de facsímiles y datos. También se utiliza dicha banda para radiotransmisión, comunicaciones móviles y transmisiones en frecuencia estándar. En la banda existen varias localizaciones (intervalos) para aficionados, y la región de 21.85 a 21.87 MHz está reservada, por tratados internacionales, para la radioastronomía<sup>1</sup>.

La propagación en la banda HF es esencialmente por medio de ondas terrestres para cortas distancias más allá del horizonte (más lejos durante el día y en las

frecuencias menores) y es ionosférica para enlaces de largo alcance, especialmente por la noche. Debido a las variaciones diurnas y de otro tipo, a menudo se establece más de una frecuencia para un solo enlace, de forma que sea posible mantener la comunicación cuando varían las condiciones. En general, se emplean dos frecuencias en un día determinado: una durante el día y otra durante la noche, pero el par de frecuencias podría cambiar de estación a estación y de un año a otro, dependiendo de las predicciones ionosféricas para el momento.

Los enlaces en HF suelen adolecer de varias condiciones graves de ruido, esencialmente de la atmósfera. La actividad de las tormentas eléctricas se presenta de forma constante en la atmósfera terrestre, y el ruido impulsivo generado por los relámpagos se propaga a la vecindad del enlace a través de la ionosfera. La densidad espectral de potencia del ruido atmosférico es inversamente proporcional a la frecuencia en casi todo el intervalo de radiofrecuencias. En la banda HF su nivel es sustancialmente mayor que el de las fuentes de ruido galáctico y otras, con la posible excepción del intenso ruido creado por el ser humano en zonas urbanas<sup>8</sup>. Es más dominante en regiones tropicales que en los polos, y varía con la estación, en ambos casos siguiendo la incidencia de las tormentas eléctricas<sup>4</sup>. Por tanto, es necesario tomar en cuenta estos factores en el cálculo de la potencia requerida para transmisión, a fin de dotar al receptor de una razón suficiente de señal sobre ruido (cap. 47).

Sin embargo, la restricción más seria en el diseño de enlaces HF es el grado de distorsión de trayectoria múltiple que sucede a menudo. La presencia de varias capas ionosféricas a diferentes altitudes hace posible la recepción de varias señales reflejadas con retardos considerablemente distintos. Además, cada reflexión de una capa puede estar sometida a un menor intervalo de retardos debido a no homogeneidades en la capa. La dispersión de retardos global puede alcanzar hasta 8 ms, y no es raro que sea del orden de 1 ms. A menos que se utilicen complicados sistemas autoadaptables de compensación del canal, esto restringirá gravemente el intervalo de frecuencias (ancho de banda) que es posible transmitir sin distorsión en cualquier esquema analógico de modulación, y, de forma equivalente, la velocidad de transmisión de los símbolos en un esquema digital.

El ancho de banda de coherencia de un canal es la mínima separación de frecuencias entre dos señales que no están correlacionadas estadísticamente. En el caso de un canal de trayectoria múltiple, el ancho de banda de coherencia es inversamente proporcional a la dispersión en tiempo de la respuesta de pulso del canal. Para una dispersión en tiempo de 1 ms, sería del orden de 1 kHz, y el valor exacto depende de la forma de la respuesta de pulso del canal. Un enlace en HF sin compensación en el que sólo se utilizara modulación analógica sufriría una seria distorsión si su ancho de banda fuese bastante superior al ancho de banda de coherencia del canal, por lo que se acostumbra a elegir el esquema de modulación de banda más estrecha posible.

Esto ayuda a explicar, por ejemplo, la popularidad de la AM de una sola banda lateral (SSB, de *single sideband*) para enlaces de habla de HF; el habla inteligible puede transmitirse con un ancho de banda aproximado de 3 kHz.

En una transmisión digital, la dispersión de retardos en sí es una medida más útil de las características del canal. La dispersión de los tiempos de llegada de la señal determina la medida en que un símbolo dado transmitido se superpone a sus vecinos en uno de los dos lados (interferencia entre símbolos). Por consiguiente, la velocidad de transferencia de los símbolos (número de símbolos por segundo) de un enlace digital en HF sin compensación no puede ser mayor que una parte del recíproco de la dispersión de retardos, lo que restringe la mayor parte de los enlaces en HF a menos de 1 000 símbolos por segundo. Para aplicaciones en las que se requieren velocidades de transferencia de datos mayores de 1 kb/s (kilobits por segundo), entonces es necesaria la codificación en la fuente para que cada símbolo transmitido represente más de un bit de información (secs. 47.3 y 48.5). La extensión de la velocidad de transferencia de bits por medio de codificación en la fuente es una técnica de alguna forma limitada, ya que el ancho de banda resultante aumenta de modo exponencial a medida que el número de bits por símbolo aumenta linealmente<sup>9</sup>. Una de las técnicas más populares es la QPSK (de *quadrature phase shift keying*), ya que duplica la velocidad de transferencia de bits sin incrementar el ancho de banda (secs. 48.1, 48.5 y 48.7).

Un método más afortunado para la comunicación en HF a gran velocidad de transferencia de datos consiste en el empleo de técnicas de compensación del canal junto con el sondeo del canal<sup>10</sup>. Si se conoce la sucesión de retardos del canal, es posible introducir un mecanismo de filtro en el receptor para compensar la distorsión de trayectoria múltiple<sup>11</sup>. Dado que las características del canal cambian con el tiempo, también es necesario introducir un mecanismo para "sondear" de vez en cuando el canal y ajustar de manera acorde el filtro de compensación.

El sondeo ionosférico puede efectuarse como parte integral de la técnica de señalización o bien como una operación del todo independiente.

**Sondeo independiente.** Las técnicas de sondeo independiente incluyen la transmisión de una señal especialmente diseñada (un pulso, un chirrido o una sucesión pseudoaleatoria) a la ionosfera y la observación de la respuesta.

**Sondeo vertical (ionosondeo).** Es una técnica que se aplica de forma regular como parte del procedimiento de predicción ionosférica.

**Sondeo a ángulos oblicuos.** Es posible si se cuenta con un receptor colocado idóneamente, así como con una red de sincronización (o si se utilizan relojes atómicos en el transmisor y el receptor); o si existe suficiente energía retrodispersada por recibir en el sitio de transmisión.

**Sondeo integral del canal.** Puede efectuarse por medio de la intercalación periódica de una señal especial de sondeo dentro de la señal del mensaje. Tales señales suelen denominarse sucesiones de aprendizaje, y pueden utilizarse también para sincronización.

Otra técnica consiste en multiplicar cada bit del mensaje por un periodo de una sucesión pseudoaleatoria de bits (sec. 48.5), de forma que el tiempo de resolución de la señal aumente allí donde sea posible resolver cada retraso de trayectoria individualmente (es decir, la función de autocorrelación de la sucesión pseudoaleatoria tiene un pico central más estrecho que un intervalo de retardo típico). Esta técnica se denomina RAKE ("rastrillo")<sup>11</sup>.

El empleo de receptores adaptables y de técnicas de sondeo del canal da por resultado un incremento de la velocidad de señalización en la banda de HF de por lo menos un orden de magnitud, con una mejora del rendimiento en términos de errores.

Es bien sabido que también pueden lograrse mejoras en el rendimiento a través del empleo de técnicas de diversidad (recepción múltiple), pero el canal ionosférico no se presta para tales técnicas.

**Diversidad espacial.** Es difícil de utilizar debido a que la separación requerida (de varios cientos de longitudes de onda) significa más superficie de terreno ocupado que el disponible para el receptor promedio.

**Diversidad de frecuencias.** Es factible desde el punto de vista técnico (la separación requerida de frecuencias a menudo es de sólo unos cuantos kilohertz), pero la banda HF ya se encuentra demasiado saturada y las agencias de reglamentación son reacias a dar cabida a múltiples canales para uso simultáneo en un enlace.

**Diversidad temporal.** No es factible en términos generales, debido a que el tiempo de coherencia es demasiado grande (de varios segundos incluso con desvanecimiento rápido).

#### Enlaces de tropodispersión

Los radioenlaces de dispersión troposférica (tropodispersión) operan en las bandas de VHF a UHF (30 MHz a 3 GHz) y más allá, con distancias de enlace hasta de 1 000 km. Debido a que la región de dispersión es más compacta que la ionosfera, las dispersiones de retardos de trayectoria múltiple son significativamente menores, proporcionando anchos de banda de coherencia del orden de 500 kHz que disminuyen con el aumento de la distancia. Así, son posibles mayores anchos de banda transmitidos y mayores velocidades de transferencia de datos que en la banda de HF.

Las bandas de VHF y UHF se utilizan para telefonía y telegrafía a larga distancia, especialmente en forma multicanalizada; para transmisión de televisión y de FM; para sistemas de radiolocalización y sistemas de aterrizaje por instrumentos; y para sistemas de comunicación por satélite. Además, existen asignaciones para radioastronomía, investigación espacial y para radioaficionados. El ruido atmosférico y el producido

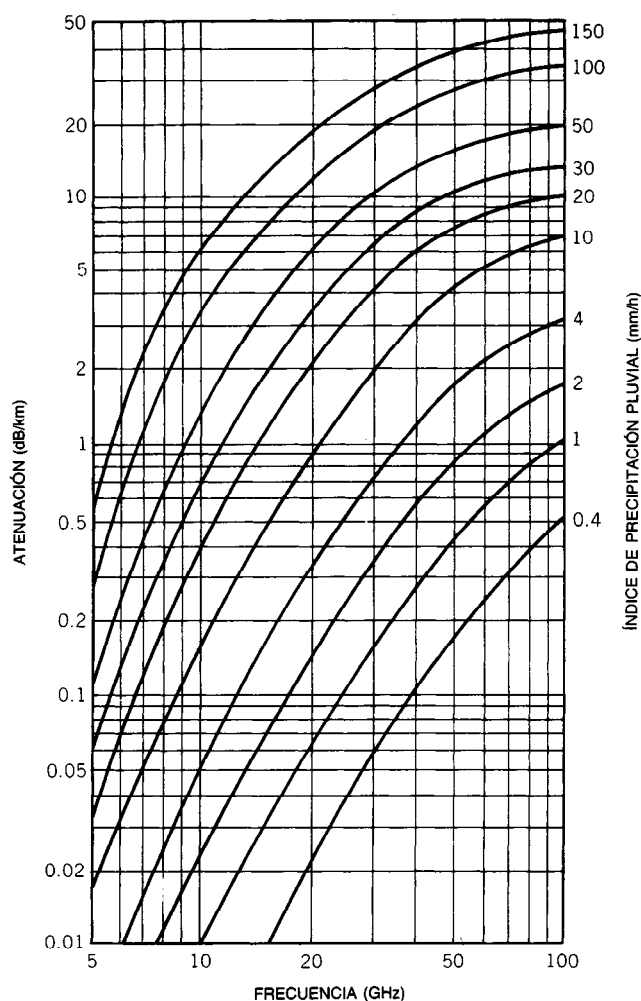
por el ser humano son significativamente menores que a bajas frecuencias y declinan inversamente con la frecuencia, y a menudo descienden por debajo del fondo galáctico en la región de 1 GHz y más<sup>12</sup>. La atenuación debida a la precipitación es despreciable, excepto en las tormentas más violentas a las frecuencias superiores (fig. 49-6).

Un beneficio más de las frecuencias superiores es la capacidad de utilizar más antenas direccionales de tamaño razonable, lo que da por resultado mejores razones de señal sobre ruido recibidas y la posibilidad de volver a utilizar la misma frecuencia para diferentes servicios en una zona geográfica dada. Además, existe a disposición más espacio espectral: 2.07 GHz de ancho de banda en comparación con los 2.7 MHz en la banda de HF. Estos factores contribuyen a la cada vez mayor obsolescencia de la banda HF.

El que el empleo del enlace troposférico sea rentable en comparación con un enlace LOS depende del tráfico de comunicación por portar y de la topología geográfica de la trayectoria del enlace. Para una longitud de trayectoria dada, un enlace LOS requiere por lo menos un repetidor para cada distancia de radio-horizonte (horizonte radioeléctrico); el enlace de dispersión troposférica ciertamente requeriría menos o tal vez ninguno. Por tanto, un enlace LOS de larga distancia costaría invariablemente más, pero estos enlaces suelen tener anchos de banda de coherencia muy amplios, por lo cual son capaces de manejar mucho más tráfico. Si existe suficiente demanda, los ingresos generados por el enlace LOS suelen compensar la desventaja original del costo de capital. El enlace de dispersión troposférica también requiere cantidades sustanciales de potencia transmitida (del orden de 1 kW), de forma que la utilización de frecuencia es más deficiente que con los enlaces LOS, debido a las restricciones de interferencia.

La tropodispersión es atractiva en zonas en las que el desarrollo del sitio de la repetidora sería exageradamente costoso, como en la selva o en regiones montañosas. Por consiguiente, los enlaces de dispersión troposférica suelen encontrarse en los países menos desarrollados del mundo. (Obsérvese que en los enlaces de una isla a otra isla más allá del horizonte *debe* utilizarse la dispersión troposférica en esta banda, ya que el empleo de repetidoras no es práctico. Las únicas opciones son el cable submarino o en enlace por satélite.)

Muchas de las técnicas para ampliar la velocidad de transmisión en los enlaces de HF se han adaptado para los enlaces de tropodispersión. Así, con frecuencia se utilizan de forma conjunta técnicas de transmisión de espectro extendido y trayectoria múltiple a fin de obtener enlaces a velocidades de megabits por segundo. Los esquemas de diversidad también son comunes; las longitudes de onda cortas permiten emplear diversidad espacial con un distanciamiento razonable de la antena (p. ej., 100 longitudes de onda se traducen en 30 m a 1 GHz). La diversidad espacial tiene la ventaja de que no se requieren potencia transmitida o ancho de banda extras, y el diseño del equipo de transmisión no es afectado por su empleo. La diversidad de frecuencias también es factible si pueden obtenerse las dos asig-



**Fig. 49-6.** Gráfica de la atenuación de las señales de microondas debido a la precipitación. Adaptado de la CCIR XIIIth Plenary Assembly, Vol. V, informe 233-3, Ginebra.

naciones de frecuencias, separadas por varias veces el ancho de banda de coherencia. Es poco probable que la diversidad temporal sea útil en la mayor parte de las aplicaciones, debido a que el tiempo de coherencia del canal, del orden de los minutos, es demasiado grande.

#### Enlaces LOS de microondas

A frecuencias de microondas ( $> 1$  GHz), las ondas de radio comienzan a presentar algo más de las propiedades de radiación óptica que a radiofrecuencias inferiores. Por tanto, el término "en línea recta" se asocia más a menudo con estas frecuencias que con cualesquiera otras. Los enlaces LOS de microondas requieren que ni la Tierra ni cualquier otro obstáculo significativo obstruyan la trayectoria entre el transmisor y el receptor, pero sería inadecuado suponer que todo lo que se requiere es el establecimiento de una trayectoria en línea recta entre ellos.

Debido a que la densidad de la atmósfera tiende a variar con la altitud, ocurre refracción y la onda "directa" se curva. El efecto es el mismo que si no existiera refracción y el valor del radio de la Tierra fuese diferente del verdadero. En condiciones normales el "radio terrestre efectivo" resultante es aproximadamente 8 500 km (casi cuatro tercios del valor medio verdadero, 6 378 km). Sin embargo, en varias condiciones atmosféricas la razón del radio efectivo entre el radio verdadero puede variar de 0.6 a 5.0. Por tanto, a fin de incluir la peor de las condiciones, es necesario incrementar el espacio libre entre la Tierra y el rayo de trayectoria recta más cercano a partir del que se haya obtenido mediante el empleo de un sencillo modelo geométrico.

En la figura 49-7 se muestra la forma de un enlace LOS. A fin de determinar si un punto P cerca de la trayectoria del rayo está lo suficientemente separado

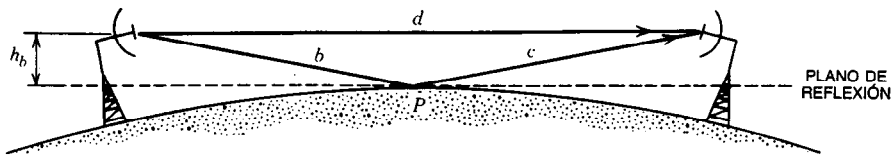


Fig. 49-7. Interferencia entre las ondas dirigida y reflejada por la Tierra en un enlace de microondas.

de éste, se denota como  $b$  a la distancia del transmisor a  $P$ , como  $c$  a la distancia de  $P$  al receptor, y como  $d$  a la distancia del transmisor al receptor. Entonces el radio  $r$  de la primera zona de Fresnel a partir de  $P$ , medido en un plano perpendicular al rayo, está dado por:

$$R^2 = \frac{\lambda bc}{d}$$

donde  $\lambda$  es la longitud de onda y todas las mediciones están en las mismas unidades.

Se considera que una obstrucción tiene suficiente espacio libre con respecto al rayo si se encuentra fuera de la primera zona de Fresnel, medido a partir del paso más próximo del rayo. Los cálculos resultantes pueden utilizarse para determinar la altura requerida de la antena para una distancia de enlace dada, o la máxima distancia de enlace para un par de alturas de antena dadas, cuando es conocido el perfil del terreno<sup>1</sup>.

En las trayectorias LOS también ocurre el fenómeno de trayectoria múltiple, como se describió en la sección 49.1.2. En los enlaces se emplea tanto la diversidad de frecuencias como la espacial<sup>12</sup>, algunas veces combinadas. Sin embargo, las agencias de reglamentación

limitan la disponibilidad de canales para la diversidad de frecuencias, por lo que es probable que los esquemas de diversidad espacial tiendan a predominar en el futuro.

La diversidad espacial para combatir el problema requiere una separación vertical entre las antenas de aproximadamente:

$$d_s = \frac{\lambda d}{4h_t}$$

donde  $h_t$  es la altura de la antena de transmisión por encima de un plano tangente a la Tierra en el punto de reflexión (fig. 49-7), y las unidades de todas las mediciones son las mismas.

La diversidad temporal no se emplea en los enlaces LOS de microondas, ya que el tiempo de coherencia es del orden de horas en condiciones normales.

Los enlaces LOS de microondas también adolecen de atenuación debida a la precipitación y absorción en la atmósfera por moléculas de oxígeno y agua. La absorción por moléculas de agua tiene un pico justamente después de los 2 GHz, y la de oxígeno, a los 60 GHz, por lo que estas frecuencias suelen evitarse en el diseño

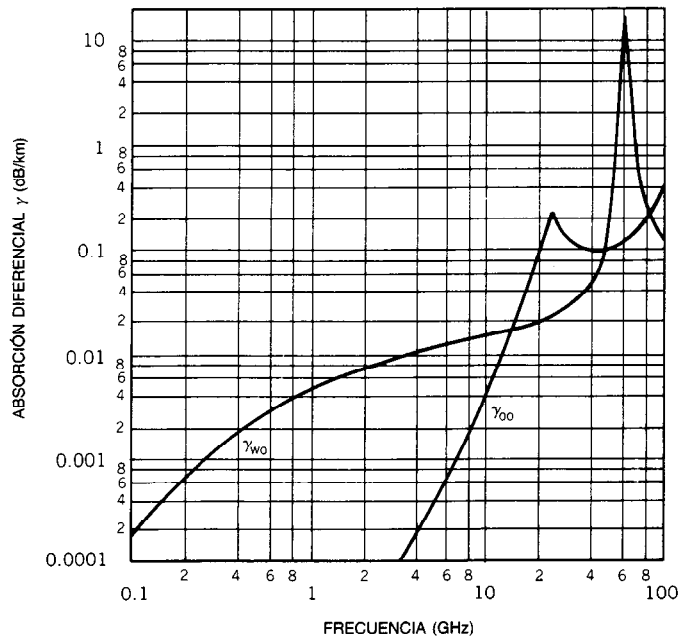


Fig. 49-8. Absorción atmosférica en función de la frecuencia. Adaptado de la CCIR XIIIth Plenary Assembly, Vol. V, informe 233-3, Ginebra.

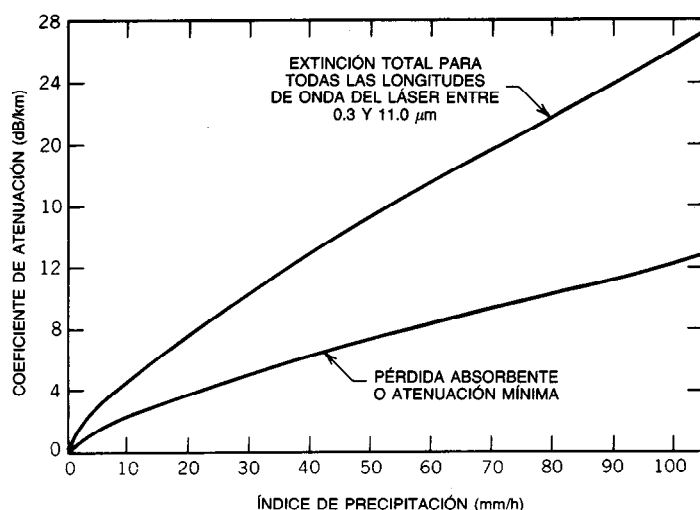


Fig. 49-9. Atenuación óptica para la precipitación en función del índice pluvial.

de enlace. En la figura 49-8 se presentan curvas de atenuación por precipitación en función de la frecuencia, con la densidad de precipitación como parámetro.

Estos enlaces tienen el mayor ancho de banda de coherencia de todos (del orden de 10 a 100 MHz), con la excepción de los enlaces ópticos, lo que permite la transmisión a velocidades hasta de 300 Mb/s. Es posible la transmisión directa de señales de televisión y grupos muy grandes de enlaces telefónicos en forma multicanalizada sin equalización. Por consiguiente, la capacidad de generación de ingresos del enlace se compara favorablemente con el gran costo de capital en la mayor parte de los casos, incluso cuando se requieren varias estaciones repetidoras para completar un enlace con gran distancia de transporte.

La generalidad de los métodos de transmisión multicanalizada que se utilizan actualmente son analógicos, y los más populares de multicanalización por división de frecuencia son el de FM-FM y el de SSB-FM (sec. 48.1). En los nuevos sistemas de construcción reciente se utilizan métodos digitales, en los que se han desarrollado sistemas de multicanalización por división tanto de frecuencia como de tiempo. Los últimos son más populares cuando se utilizan repetidoras saturadas, y parece probable que en el futuro dominen los métodos de acceso múltiple por división de tiempo (TDMA, de *time division multiple access*).

#### Enlaces LOS ópticos

Estos enlaces operan a frecuencias superiores a 30 THz ( $3 \times 10^{13}$  Hz) y ofrecen el mayor ancho de banda de coherencia de cualquier banda de frecuencia (del orden de 1 a 10 GHz). Por tanto, existe el potencial para la transmisión de grandes grupos multicanalizados de señales de televisión y enormes grupos de enlaces telefónicos y datos en un solo canal (sec. 48.4). Por desgracia, los enlaces atmosféricos ópticos están restringidos a líneas rectas en el sentido más literal, se encuentran sometidos a seria atenuación en presencia

de precipitación y aerosoles atmosféricos, experimentan grave absorción por las moléculas de los gases atmosféricos, y son muy susceptibles a la refracción debida a efectos atmosféricos térmicos y turbulentos. Estos factores se combinan con las dificultades prácticas de alineamiento mecánico de las "antenas" con haces muy estrechos y con el éxito técnico y bajo costo de los enlaces de fibra de vidrio, para hacer del enlace óptico LOS una opción muy poco atractiva en casi todos los casos, excepto cuando se trata de los enlaces más cortos.

Consecuentemente, es posible esperar que tales enlaces se utilicen sólo para el tránsito de alta densidad con corta distancia de transporte (de algunos cientos de metros) en zonas urbanas y en zonas en las que la instalación de enlaces de fibra es costosa o imposible, como con islas cercanas o satélites. (Probablemente la aplicación más promisorio de los enlaces ópticos LOS es en el espacio exterior, donde no se encuentran estos problemas.)

La absorción molecular ocurre en puntos fijos del espectro debido a la presencia de gases y vapores atmosféricos. La estructura de la absorción es demasiado detallada para ser presentada en una gráfica que cubra todo el espectro del infrarrojo al ultravioleta; por ejemplo, en la tabla 49-2 se muestran algunas fuentes importantes de láser de infrarrojo que son fuertemente absorbidas por el agua atmosférica<sup>13</sup>. Es posible elaborar tablas semejantes para muchas otras moléculas atmosféricas, incluyendo algunos absorbentes fuertes que se presentan sólo en cantidades infinitesimales. De este modo, la atmósfera se caracteriza por muchas "ventanas" en el espectro, con frecuencia muy estrechas, alternando con regiones casi opacas en las que la propagación es casi imposible. La elección de una frecuencia de operación, y por tanto de la fuente del láser, es de importancia crítica.

También ocurre dispersión molecular, dependiendo de la densidad del gas en la trayectoria de la propaga-

ción y por tanto de la altitud. Este fenómeno se denomina dispersión de Rayleigh (que es el fenómeno por el cual el color del cielo es azul), y la atenuación resultante es proporcional a la frecuencia elevada a la cuarta potencia.

**Tabla 49-2. Líneas de láser en la región infrarroja fuertemente absorbidas por moléculas de vapor de agua**

Fuente de láser	Longitud de onda ( $\mu\text{m}$ )
Kriptón atómico	1.7843
Kriptón atómico	1.9211
$\text{Tm}^{+3}\text{-CaWO}_4$	1.911
$\text{Tm}^{+3}\text{-CaWO}_4$	1.916
$\text{U}^{+3}\text{-SrF}_2$	2.472
$\text{U}^{+3}\text{-CaF}_2$	2.511
Kriptón atómico	2.5234
$\text{U}^{+3}\text{-BaF}_2$	2.556
$\text{U}^{+3}\text{-CaF}_2$	2.613
CO	5.2 a 7
Cesio	7.1821
Neón atómico	18.3040
Neón atómico	20.351

El efecto de los aerosoles es el mecanismo de atenuación más variable, ya que depende del tipo y densidad de tales materiales en la propagación y por tanto varía con la localización, altitud, patrones de clima, actividad industrial y volcánica, y otros factores. Por ejemplo, en la figura 49-9 se muestra una gráfica del coeficiente (factor) de atenuación en función de la intensidad de la precipitación.

Además de producir los efectos de absorción, la turbulencia atmosférica rompe el frente de ondas plano del haz óptico y provoca pérdidas debidas a dispersión y falta de coherencia. La distancia de coherencia de la onda recibida suele ser menor que el diámetro de la abertura del receptor, lo cual da por resultado una

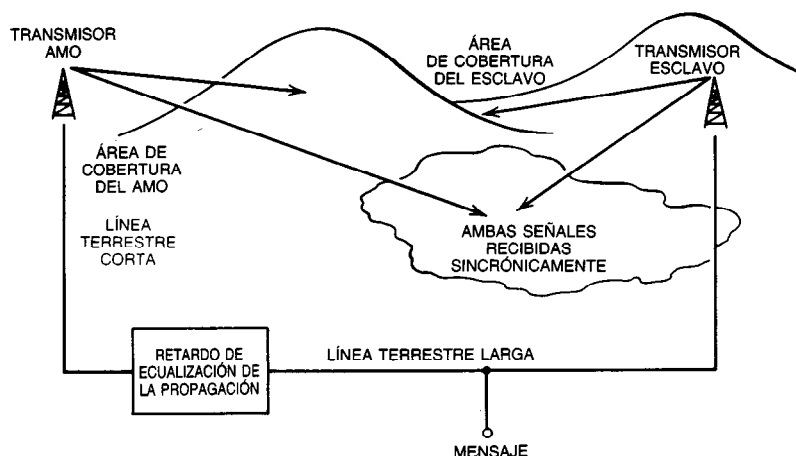
“abertura efectiva” que puede ser significativamente menor que el tamaño físico<sup>14</sup>. Por tanto, no es posible obtener toda la ganancia teórica de potencia de la abertura del receptor si la abertura es grande en comparación con la distancia de coherencia del canal (factor que también limita el tamaño útil de los telescopios astronómicos ópticos para ser utilizados en la Tierra). Si la abertura del receptor contiene varias zonas con intensidades de señal estadísticamente independientes (modos ópticos), las potencias resultantes se suman, en vez de las amplitudes resultantes. Esto es el equivalente de la diversidad espacial en los sistemas de radio, y ayuda a eliminar el “centelleo” debido a la turbulencia<sup>15</sup>.

Los métodos de modulación pueden consultarse en la sección sobre sistemas de comunicaciones con fibras ópticas (48.4).

#### 49.1.4 Radioenlaces móviles

Existen asignaciones de frecuencias para radios móviles en casi todas las bandas del espectro electromagnético regulado. La banda elegida depende de la aplicación: en los sistemas marinos y submarinos de larga distancia se emplean las bandas HF y menores; en los servicios marítimos se utiliza la VHF para servicio de corta distancia; en los servicios de aeronáutica se utiliza la banda HF y más altas para la comunicación y las bandas MF y LF para la navegación; y en los sistemas terrestres móviles se emplean las bandas VHF y UHF en casi todos los casos. En la tabla 49-3 se muestran las asignaciones más importantes para radios móviles en Estados Unidos. Las comunicaciones móviles son quizá la más desafiante de todas las aplicaciones, principalmente debido a que el canal varía con el tiempo en virtud del movimiento del vehículo, y es de naturaleza incontrolable debido al carácter aleatorio de la localización del vehículo.

En la mayor parte de los casos de interés, el canal de radio móvil se comporta como el canal de dispersión local descrito en la sección sobre propagación. En mu-



**Fig. 49-10.** Principio de la transmisión simultánea.

chos sentidos este canal es el "peor caso", de forma que es probable que un diseño de sistema que funcione satisfactoriamente en tal canal también lo haga en casi cualquier otra situación.

La mayor parte de los servicios móviles desplegados en aplicaciones civiles hacen uso de técnicas de modulación analógica, y más a menudo de la modulación de frecuencia de banda estrecha. Para la denominada banda civil se utiliza una doble banda lateral DSB (de *double sideband*) de AM en la región de los 27 MHz, y se está desplegando un AM-SSB en algunas bandas congestionadas en un esfuerzo por mejorar la disponibilidad del sistema. Se están comenzando a utilizar las técnicas digitales, pero las restricciones legales y tecnológicas con frecuencia las han limitado a señales de manipulación de tono en la banda base que sencillamente se modulan de manera analógica en una portadora. Esto restringe seriamente la máxima velocidad de transferencia de bits a algunos kilobits por segundo, aunque sigue siendo útil en aplicaciones como terminales de teclado remotas para el acceso a bases de datos. El uso común y amplio de radios móviles digitales de velocidad media y alta aún parece algo lejano. Ya que se ha demostrado su factibilidad técnica y debido a que su costo no es elevado, actualmente el principal obstáculo es el proceso de reglamentación.

**Tabla 49-3. Algunas asignaciones importantes de radio móvil público en Estados Unidos<sup>a</sup>**

35.19	a	35.69 MHz
43.19	a	43.69 MHz
152.0	a	152.255 MHz
152.495	a	152.855 MHz
157.755	a	158.115 MHz
158.475	a	158.715 MHz
454.0	a	455.0 MHz
459.0	a	460.0 MHz
470.0	a	512.0 MHz
806.0	a	821.0 MHz
825.0	a	845.0 MHz <sup>b</sup>
851.0	a	866.0 MHz
870.0	a	890.0 MHz <sup>b</sup>

<sup>a</sup> Consúltense una lista completa en la referencia 1.

<sup>b</sup> Asignaciones celulares

### Servicios móviles terrestres privados

Los servicios de radio móviles terrestres (LMR, de *land mobile radio*) privados abarcan todos los servicios en los que se utilizan asignaciones fijas para un solo propósito y un solo usuario o grupo de usuarios. Estos servicios pueden ser para usos de negocios, transporte, industriales, de seguridad pública, y públicos en los que un usuario dado necesita utilizar de forma exclusiva un canal o conjunto de canales en sentido local.

En la actualidad, los métodos de modulación para tales aplicaciones son casi exclusivamente de FM con modulación de tono en la banda base para enlaces digitales. En la generalidad de los casos el ancho de ban-

da transmitido es menor que el ancho de banda de coherencia del canal, por lo que se considera que estas transmisiones adolecen de "desvanecimiento plano"; es decir, desvanecimiento no selectivo de frecuencia, con estadísticos de intensidad de campo de Rayleigh. El efecto del desvanecimiento no suele tomarse en cuenta para el diseño del receptor, excepto por el uso ocasional de la diversidad espacial. En la potencia nominal del transmisor se incluyen márgenes de desvanecimiento del orden de 10 dB, en un intento por reducir la probabilidad de paralización cuando el receptor se mueve alrededor de la zona de servicio. Si ocurre una paralización, existe una gran probabilidad de que se logre el funcionamiento satisfactorio alejando la antena del receptor apenas unas cuantas longitudes de onda del "punto muerto", de forma que rara vez se requiere el empleo de técnicas complicadas.

La mayor parte de los enlaces LMR de este tipo vinculan directamente una unidad móvil y alguna estación base ubicada centralmente, o dos unidades móviles. Ya que la zona de servicio a menudo es muy grande (tal vez toda una ciudad), las estaciones base suelen ubicarse lo más elevadas posible en un lugar central cuidadosamente elegido, y las potencias de transmisión son lo suficientemente altas a fin de proporcionar cobertura más allá de los límites de la zona de servicio. Si la potencia del transmisor es inadecuada para la comunicación entre dos unidades móviles directamente, la estación base puede actuar como retransmisor ya sea "parchando" los dos enlaces o retransmitiendo verbalmente el mensaje.

Cuando se requiere amplia cobertura, especialmente en terrenos montañosos, se utiliza la técnica conocida como transmisión simultánea (*simulcasting*). Esta técnica implica la transmisión de la misma portadora y modulación por dos o más estaciones base ubicadas en sitios diferentes, y da por resultado una señal recibida que se parece bastante a una señal con serio efecto de trayectoria múltiple (fig. 49-10). Debe tenerse el cuidado de sincronizar las portadoras y compensar los retardos de propagación en la transmisión múltiple, o se recibirá una señal distorsionada.

### Sistemas LMR de portadora común

Radiotelefonía móvil, localización de personas (buscapersonas) y algunos sistemas privados de comunicación son proporcionados por compañías que tienen suficientes asignaciones de canales para ofrecer un servicio accesible a una gran población de suscriptores, cada uno con un pequeño tiempo promedio de llamadas diarias (5% o menos). Estas compañías se denominan portadoras comunes y a menudo se identifican con compañías telefónicas locales con servicio alambrado (p. ej., el sistema Bell). La utilización de cada canal está determinada por principios parecidos a una línea o "tronco" de una red telefónica de larga distancia. Si existe a disposición un canal, es posible completar la llamada: en caso contrario, ocurre "bloqueo" y se escucha una señal de ocupado.

En la generalidad de los servicios actuales de portadora común se emplea una sola estación base por zona de servicio, de forma que cada canal puede utilizarse



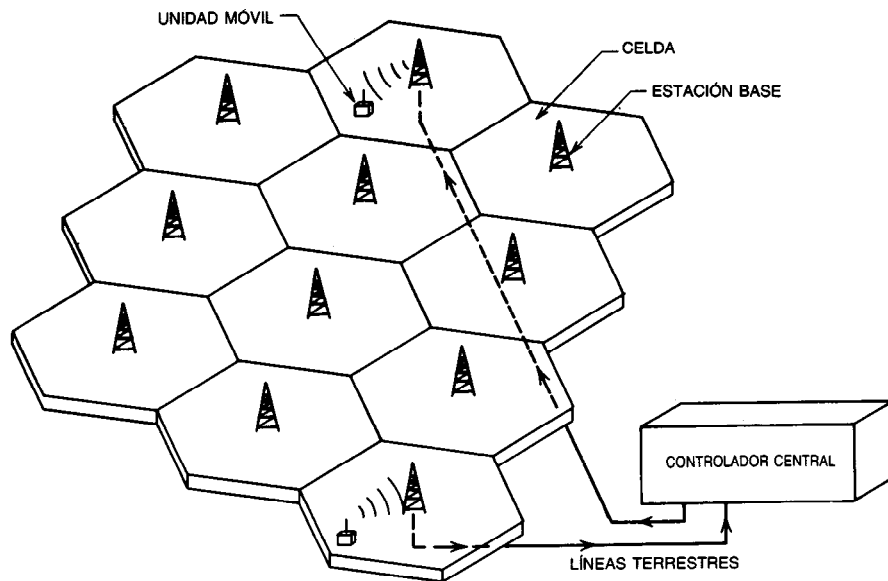


Fig. 49-11. Concepto del sistema celular de radios móviles terrestres (LMR).

solamente una vez en cualquier instante dado en la zona de servicio (y para alguna distancia más allá de la zona de servicio, siempre que se requiera la transmisión de grandes potencias). Por consiguiente, el número de llamadas simultáneas por zona de servicio está limitado a la mitad del número disponible de canales, ya que se requieren dos canales para una conversación telefónica dúplex. Además, no todos los canales de las portadoras comunes pueden asignarse a una sola zona de servicio, debido a que ciudades vecinas que requieran servicio pueden estar demasiado cerca entre sí para permitir la reutilización de los mismos canales. El resultado ha sido una grave congestión en muchas zonas metropolitanas y una gran lista de usuarios en espera de una licencia de operación. Actualmente se está poniendo en práctica el empleo de formatos celulares a fin de solucionar el congestionamiento.

La modulación de frecuencias de banda estrecha, con separaciones de canales en la región de 15 a 30 kHz, es la norma para la transmisión del habla en casi todos los sistemas LMR de portadora común. La diversidad espacial se ofrece ocasionalmente como una opción del equipo de recepción para reducir los efectos del desvanecimiento<sup>2</sup>. La transmisión de datos, que incluye los canales de control utilizados para el establecimiento de llamadas en un sistema de radiotelefonía, emplea modulación por manipulación de tono en la banda base. En los canales de datos suele utilizarse diversidad temporal, ya que la velocidad de transferencia de datos no es elevada y el tiempo de coherencia para los canales de radio móvil es típicamente 0.1 s para unidades móviles en una parada e inferior cuando la unidad se encuentra en movimiento.

Recientemente se ha propuesto el empleo de la AM-SSB, tal vez incluyendo técnicas de compresión/expansión digitales, como solución provisional para el

problema del congestionamiento<sup>17</sup>. Se ha propuesto el empleo de una separación de 5 kHz en los canales disponibles, pero aún queda por demostrar si el funcionamiento en presencia de desvanecimiento, interferencia de canal común e interferencia de canales adyacentes será satisfactorio, si la distorsión de la señal provocada por la compresión/expansión de amplitud y frecuencia será aceptable, y si es posible producir de forma efectiva en cuanto a costo el equipo con las tolerancias de frecuencia necesariamente reducidas<sup>18</sup>. Si estos problemas pueden ser resueltos de forma satisfactoria, entonces la idea SSB puede tener un futuro a corto plazo. Sin embargo, el problema de la transmisión digital en tales enlaces, su compatibilidad con un creciente servicio telefónico alambreado digital, y su idoneidad para ser utilizada en esquemas celulares parecen poco probables de resolver.

#### Sistemas LMR celulares

Un sistema LMR celular es un esquema en el cual el área de servicio se divide en "zonas de miniservicio" más pequeñas, a cada una de las cuales se dota de su propia estación base. Las celdas suelen representarse como cuadrados o como rectángulos, pero en una zona real de servicio las fronteras de las celdas se encuentran localizadas allí donde la intensidad de las señales de estaciones base es aproximadamente la misma. La comunicación entre una estación (o unidad) móvil y una estación base se establece por medio de un radioenlace corto entre la estación base con la mejor razón de señal sobre ruido medida en la estación móvil, que *casi seguramente* será la más cercana. La comunicación entre dos estaciones móviles se establece mediante dos de tales radioenlaces, uno en la celda de cada estación, acoplados con una línea de tierra o con otro enlace no radiativo entre las dos estaciones base (fig. 49-11). Por

consiguiente, la potencia transmitida requerida es sólo la necesaria para el enlace más largo, cuya distancia no excede el "radio" de cada celda. Cuando una unidad móvil se desplaza de una celda a otra, ambos enlaces del mensaje son recanalizados por medio de la estación base de la nueva celda de la estación móvil. (Esto implica que el controlador central debe mantener el seguimiento de la ubicación de todas las unidades móviles<sup>20</sup>.) Esta característica proporciona un medio sencillo para limitar y controlar la interferencia entre usuarios que ocupan la misma porción del espectro de forma simultánea, y por tanto abre el camino para un uso más eficiente del espectro.

A fin de lograr la meta de eficiencia en un sistema LMR celular se han propuesto dos formas diferentes, que se describen a continuación.

**Esquemas de reutilización de frecuencias de banda estrecha.** En estos esquemas se utiliza FM de banda estrecha o algún otro tipo de modulación en el cual el ancho de banda transmitido no excede en gran medida del ancho de banda del mensaje. Cada par de canales se asigna a más de una celda en la zona de servicio, pero celdas con la misma asignación se encuentran separadas por "anillos" de celdas que tienen asignaciones diferentes; véase la figura 49-12. Las unidades móviles deben utilizar los canales que les ha asignado (por medio de un canal de control privado) el sistema de control central. Cuando una unidad móvil se desplaza de una celda a otra, el controlador central emite una nueva asignación de canal, de forma que no se violen las nuevas asignaciones de celda. Al mismo tiempo, las señales de la llamada se recanalizan. Por tanto, en una llamada completa podrían utilizarse varias estaciones base y canales en instantes diferentes a medida que la unidad móvil se desplaza. El menor grupo de celdas

que comparten todo el conjunto de canales disponibles se denomina "cúmulo"; en la figura 49-12 cada cúmulo contiene siete celdas.

El número de celdas en un cúmulo es determinado por la distancia de reutilización del sistema, que es la mínima asignación de canal. Un sistema compuesto de cúmulos de  $N$  celdas hexagonales tiene una distancia de reutilización de  $R\sqrt{3N}$ , donde  $R$  es el "radio" de la celda; es decir, la distancia del centro a cualquiera de sus vértices. Por tanto, la distancia de reutilización de la figura 49-12 es  $\sqrt{21} = 4.58$  radios de la celda.

La intensidad de campo en una ciudad tiende a decaer con mayor rapidez de la que predecía la ley del cuadrado inverso para la propagación en el espacio libre; la ecuación de intensidad de campo es con mayor frecuencia de la forma:

$$P_r = KP_t d^{-\alpha}$$

donde  $P_t$  = potencia transmitida

$P_r$  = potencia recibida *promedio* (en la que se han eliminado los efectos del desvanecimiento)

$d$  = distancia a la estación base

$K$  = constante que depende del diseño de la antena, frecuencia de operación y factores ambientales

El exponente  $\alpha$  está en la región de 3.0 a 4.0, y el valor más bajo corresponde a las distancias más cortas (es decir, a las celdas más pequeñas). Un valor común es aproximadamente 3.5.

Debido a esta característica de propagación, la razón de señal sobre interferencia (SIR, de *signal-to-interference ratio*) promedio para una estación base que recibe señales de una unidad móvil en su propia celda a la distancia  $D$ , con interferencia de canal común

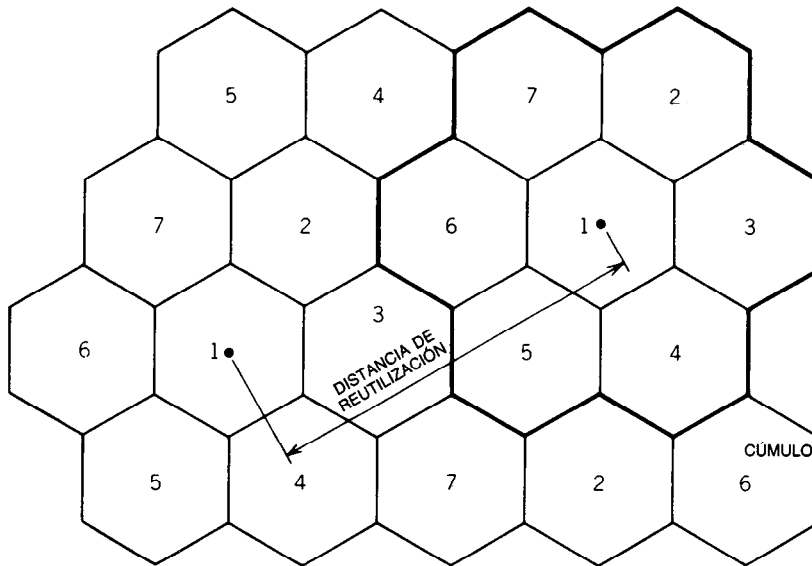


Fig. 49-12. Sistema de reutilización de frecuencias de banda estrecha con siete celdas por cúmulo. Los números se refieren al subconjunto de canales asignados a la celda.

de una unidad móvil en una celda cercana a una distancia  $d$ , está dada sencillamente en decibeles como  $SIR = 10\alpha \log(d/D)$  dB.

El punto importante que debe observarse con respecto a esta ecuación es que debido a que la razón de distancias es independiente del tamaño de la celda, sólo la forma de la celda determina las características de la interferencia. Si  $\alpha$  se mantiene fijo, el volumen de tráfico operado por una celda dada es independiente de su tamaño. Por consiguiente, es posible satisfacer un incremento en la demanda mediante la reducción de la celda, en vez de asignar más canales al sistema.

La solución intermedia (de compromiso) esencial del diseño de tal sistema se halla entre la eficiencia espectral y la confiabilidad de la comunicación. A fin de obtener un sistema eficiente se requiere el menor tamaño  $N$  del cúmulo, de forma que cada celda cuente con el mayor número posible de canales, pero la razón de señal sobre interferencia es más deficiente para  $N$  pequeña, debido a que la distancia de reutilización es menor. Por tanto, el tamaño del cúmulo es un parámetro crítico en el diseño, así como también lo es la elección de los métodos de modulación: un sistema con alguna ganancia de procesamiento (razón entre la SNR de salida y de entrada) permitirá el empleo de menores distancias de reutilización.

Para ilustrar la conexión entre el tamaño  $N$  del cúmulo, la distancia  $T$  de reutilización y el factor  $\alpha$  de propagación, considérese el siguiente ejemplo simplificado. Un sistema de siete celdas por cúmulo, con sólo dos cúmulos, funciona en una ciudad con  $\alpha = 4.0$ . Una unidad móvil situada en el borde de su celda, sobre una línea que une su estación base con la otra estación base que utiliza los mismos canales, recibe señales de ambas estaciones base; véase la figura 49-13. La razón promedio de potencia deseada sobre potencia necesaria para interferir la señal (SIR) es  $10\alpha \log(T - 1) = 40 \log(3.58) = 22.2$  dB, aproximadamente. Si se desplegara un sistema parecido en una ciudad con  $\alpha = 3.0$ , la

misma unidad móvil experimentaría una SIR promedio de 16.6 dB. Para restaurar el funcionamiento al mismo nivel que el del primer sistema, se requeriría un tamaño de cúmulo de aproximadamente 14 celdas, que en el ejemplo es todo el sistema. Por consiguiente, no habría reutilización de canales, y la capacidad de tráfico del segundo sistema sería la mitad del primero. Esto ilustra la sensibilidad respecto al parámetro  $\alpha$  del diseño de reutilización de canales de banda estrecha, que desafortunadamente es ambiental y por tanto se encuentra fuera del control del ingeniero.

El sistema Bell ha establecido en Chicago un sistema de demostración del tipo descrito<sup>19</sup>, donde se utilizan los parámetros que se proporcionan en la tabla 49-4.

**LMR celular de espectro extendido.** Es un esquema en el cual el espectro disponible se divide en sólo dos "canales" muy amplios, uno para la transmisión estación base-unidad móvil, y el otro para la transmisión unidad móvil-estación base. Todos los usuarios del sistema ocupan por completo la parte apropiada del espectro durante la llamada. Las diferentes señales se distinguen asignando a cada usuario del sistema una serie de identificación distinta, o "clave" (tales sistemas se denominan CDMA, de *code division multiple access*).

La clave de identificación se utiliza como función de dispersión para las transmisiones de espectro extendido (sec. 48.5). Cualquiera de estas funciones que sea una elección idónea para su empleo en aplicaciones CDMA también debe ser adecuada para el esquema, aunque, en dos esquemas propuestos por separado en publicaciones especializadas, se han mencionado como candidatas adecuadas funciones de saltos en frecuencia<sup>21, 22</sup>. La elevada dimensionalidad de las señales con saltos en frecuencia hace posible la asignación de una clave de identificación única a cada usuario del sistema, permitiendo de esta forma la comunicación privada y simplificando el procedimiento de direccionamiento. Sin embargo, esta característica no forma

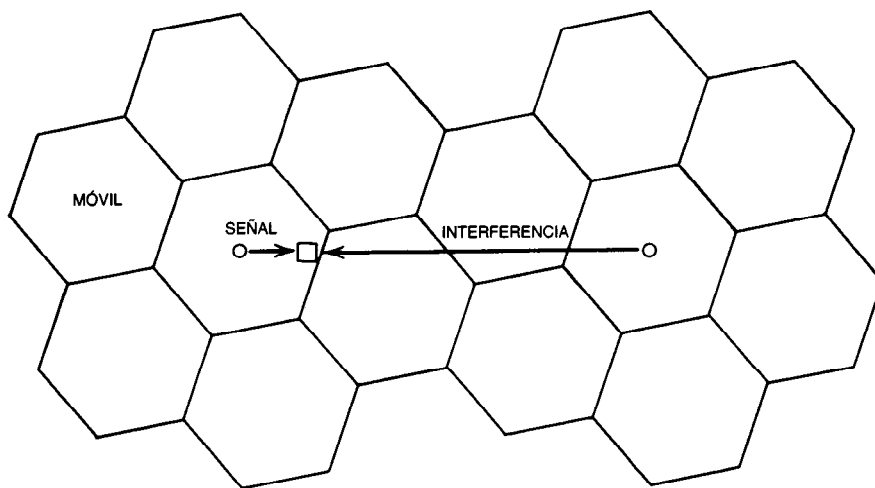


Fig. 49-13. Forma del peor caso de interferencia para un sistema celular de radio móvil terrestre (LMR).

parte inseparable del esquema, y sería satisfactorio un menor conjunto de señales, de modo que solamente existiera una clave para todos los usuarios *simultáneos* (un número mucho más pequeño que el de todos los usuarios del sistema), si cada receptor fuese capaz de recibir instrucciones para utilizar cualquier serie.

En vez de asignar sólo una parte de la banda disponible a cada celda, en el sistema de espectro extendido se utiliza toda la banda en cada celda. Las interferencias inevitables se controlan de dos formas:

- **Primero.** La potencia transmitida por cada unidad móvil se controla dinámicamente, de modo que la potencia recibida en cualquier estación base sea la misma para toda unidad móvil en tal celda.
- **Segundo.** El diseño del conjunto de claves de identificación es tal que la correlación cruzada entre cualquier par de series (claves) en el conjunto es uniformemente pequeña.

Un receptor de espectro extendido "sintonizado" con tal señal puede funcionar satisfactoriamente con una razón de señal sobre interferencia mucho menor que la unidad; de hecho, si la SIR umbral de operación del receptor es  $1/M$ , donde  $M$  es un entero positivo, entonces el sistema puede funcionar con un nivel de interferencia que sea equivalente al de  $M$  usuarios en cada estación base. Como en el sistema de banda estrecha, este número es independiente del tamaño de las celdas si el factor de propagación  $\alpha$  se mantiene fijo.

La principal ventaja que se obtiene al utilizar el espectro extendido en radios móviles es su insensibilidad a los parámetros ambientales. En términos generales, esto incluye la capacidad inherente de las señales de espectro extendido de funcionar bien en presencia de desvanecimiento. Dado que la señal transmitida se extiende sobre varios anchos de banda de coherencia del canal, cuenta con diversidad de frecuencia integrada debido a que las diferentes componentes espectrales se

desvanecen de forma independiente. En el contexto de un esquema celular, el sistema de espectro extendido también es insensible al parámetro de propagación  $\alpha$ , que resulta crítico en esquemas de banda estrecha. Por ejemplo, la capacidad de tráfico de un sistema sencillo de 14 celdas, como el que se analizó en la sección sobre banda estrecha, variaría en menos del 30% a medida que  $\alpha$  cambiara de 4.0 a 3.0, en contraste con el 100% de variación del sistema de banda estrecha. Así, podría esperarse que el sistema de espectro extendido funcione mejor en terreno difícil que el esquema de banda estrecha.

También se ha afirmado que con los esquemas LMR celulares de espectro extendido se hará uso más eficiente del espectro que con su contraparte de banda estrecha. Es difícil predecirlo; la ventaja puede ser solamente pequeña con esquemas de celdas grandes en los cuales  $\alpha$  puede controlarse con facilidad y sea bastante grande. Sin embargo, parece que cuando se hagan necesarios los esquemas de celdas pequeñas (de alta densidad), el sistema de espectro extendido será mejor en este sentido.

#### REFERENCIAS BIBLIOGRÁFICAS

1. International Telephone and Telegraph Corporation, *Reference Data for Radio Engineers*, 6.<sup>a</sup> ed., Howard W. Sams, Indianapolis.
2. W. C. Jakes, Jr., ed., *Microwave Mobile Communications*, Wiley, Nueva York.
3. A. D. Watt, *VLF Radio Engineering*, Pergamon, Elmsford, NY.
4. K. Davies, *Ionospheric Radio Waves*, U.S. Department of Commerce, National Bureau of Standards, Boulder, CO.
5. M. Leftin, ed., *Ionospheric Predictions*, U.S. Department of Commerce, Institute for Telecommunications Sciences, Boulder, CO.
6. William C. Y. Lee, *Mobile Communications Engineering*, McGraw-Hill, Nueva York.

**Tabla 49-4. Parámetros de operación para la demostración del sistema de radio móvil celular**

Tamaño del cúmulo $N$	12
Frecuencias de operación	825-890 MHz
Método de modulación del habla	FM
Separación entre canales	30 kHz
Número de canales	800 dúplex
Probabilidad de bloqueo	0.02
Método de modulación de los datos	Manipulación por desplazamiento de frecuencia con separación de 8 kHz
Velocidad de transferencia de los datos	10 kb/s
Codificación de los datos	Manchester
Corrección de errores	Código BCH (40, 28)
Diversidad temporal de los datos	5 repeticiones
Tiempo de espera en diversidad	10 ms
Diversidad espacial	Opcional

Fuente: Número especial del *Bell System Technical Journal*<sup>19</sup>, con permiso.

7. A. V. Venetsanopoulos, "Modelling of the Sea Surface Scattering Channel and Undersea Communications", en J. Skwartzynski, ed., *Communication Systems and Random Process Theory*, Sijthoff & Noordhoff, Amsterdam.
8. E. N. Skomal, *Man-Made Radio Noise*, Van Nostrand Reinhold, Nueva York.
9. J. M. Wozencraft e I. M. Jacobs, *Principles of Communication engineering*, Wiley, Nueva York.
10. M. Darnell, "Channel Evaluation Techniques for Dispersive Communications Paths", en J. Skwartzynski, ed., *Communication Systems and Random Process Theory*, Sijthoff & Noordhoff, Amsterdam.
11. R. Price y P. E. Green, Jr., "A Communication Technique for Multipath Channels", *Proc. IRE* 46:555-570. Reimpresión en B. Goldberg, ed., *Communication Channels: Characterization and Behavior*, IEEE Press, Nueva York.
12. K. Feher, *Digital Communications: Microwave Applications*, Prentice-Hall, Englewood Cliffs, NJ.
13. CRC, "Handbook of Lasers with Selected Data on Optical Technology", Chemical Rubber Company.
14. Grupo Asesor para la Investigación y el Desarrollo Aeroespacial, *Optical Propagation in the Atmosphere*, AGARD Conferencia n.º 183. Distribuido por National Technical Information Service, Springfield, VA.
15. R. M. Gagliardi y S. Karp, *Optical Communications*, Wiley, Nueva York.
16. V. E. Benes, *Mathematical Theory of Connecting Networks and Telephone Traffic*, Academic, Nueva York.
17. B. Lusignan, "Single-Sideband Transmission for Land Mobile Radio", *IEEE Spectr.*, pp. 33-37. Condensado del informe "Spectrum Efficient Technology for Voice Communications", rendido a la UHF Task Force of Office of Plans and Policy, Federal Communications Commission. Incluye cinta de demostración. Disponible en National Technical Information Service, Springfield, VA.
18. Informe de Electronics Industries Association TR-8 Ad Hoc Committee, Communications Division, for Spectrum-Efficient Technology. En los archivos del Land Mobile Communications Council, Washington, DC. Se publicó un resumen en la *IEEE Commun. Mag.* 17(2):25-28. En este informe se concluye que las recomendaciones de la ref. 17 fueron "infundadas".
19. "Advanced Mobile Phone Service", Special Issue of *Bell Sys. Tech. J.* 58(1).
20. E. N. Skomal, *Automatic Vehicle Location Systems*, Van Nostrand Reinhold, Nueva York.
21. G. R. Cooper y R. W. Nettleton, "A Spread-Spectrum Technique for High-Capacity Mobile Communications", *IEEE Transactions on Vehicular Technology*, VT-27(4).
22. D. J. Goodman, P. S. Henry y V. K. Prabhu, "Frequency-Hopped Multilevel FSK for Mobile Radio", *Bell Syst. Tech. J.* 59(9).

## 49.2 SISTEMAS DE RADIODIFUSIÓN

### Jefferson F. Lindsey III

El término radiodifusión estándar se refiere a la transmisión del habla y música recibidos por el público en general en la banda de frecuencias de 535 a 1 605 kHz. La modulación de amplitud se utiliza para un servicio que varía desde el necesario para las pequeñas comunidades hasta las estaciones de radiodifusión de potencia superior necesarias para zonas regionales grandes. La **zona de servicio primaria** se define como la zona en la cual las ondas terrestres no están sujetas a interferencia o desvanecimiento inconvenientes. La **zona de servicio secundaria** se refiere a una zona atendida por ondas indirectas y que no se encuentra sujeta a interferencia objetable. El término **servicio intermitente** se refiere a una zona que recibe servicio de una onda terrestre, pero que se encuentra más lejos que la zona de servicio primaria y está sometida a algo de interferencia y desvanecimiento.

#### 49.2.1 Radiodifusión estándar (modulación de amplitud)

##### Asignaciones de frecuencia

En Estados Unidos, las frecuencias portadoras para la radiodifusión estándar se encuentran en las *Rules and Regulations*, Volumen 3, Parte 73 de la Federal Communications Commission (FCC). Un total de 107 frecuencias portadoras están asignadas de los 540 a los 1 600 kHz en intervalos de 10 kHz, y la FCC requiere que cada una de ellas se desvíe no más de  $\pm 20$  Hz de la frecuencia asignada, a fin de minimizar la heterodinaación de dos o más estaciones interferentes. La modulación de portadora completa de doble banda lateral, denominada comúnmente modulación de amplitud (AM, de *amplitude modulation*), se utiliza en la radiodifusión estándar para transmisión del sonido. Las frecuencias de modulación comunes para el habla y la música varían de 50 Hz a 16 kHz.

En general, cada canal se concibe de modo que su ancho sea de 10 kHz, y por tanto la banda de frecuencia se diseña de 535 a 1 605 kHz; sin embargo, cuando la frecuencia de modulación excede los 5 kHz, el ancho de banda de radiofrecuencia del canal excede los 10 kHz y es posible que ocurra interferencia con el canal adyacente.

##### Clasificaciones de estaciones y canales

En la radiodifusión estándar (AM), los canales se clasifican conforme al intervalo de cobertura. Las estaciones de canal libre (o de onda exclusiva) se utilizan para prestar servicio a una amplia zona, y se limpian de interferencia inconveniente por medio del control de las concesiones y acuerdos internacionales. En Estados Unidos, la potencia máxima de salida de una estación de canal libre está limitada a 50 kW. En canales regionales se utiliza una potencia máxima de salida de 5 kW, y dichos canales están limitados por la intensidad de campo que se produce en otras estaciones de canales regionales. El tercer tipo de canal es el local, en el que

se utiliza una salida de potencia máxima de 1 kW durante el día y de 250 W por la noche, excepto para las estaciones del sureste de Florida, en las que la potencia está limitada a 250 W tanto durante el día como por la noche. Fuera de Estados Unidos, las potencias máximas de salida alcanzan un máximo de 2 MW.

De los 107 canales para radiodifusión estándar, 62 se denominan canales libres, y las estaciones en estas frecuencias pueden clasificarse como I-A o I-B. Para las estaciones de la clase I-A, la potencia de salida es de 50 kW y la zona de servicio primaria está libre de interferencia en el mismo y en canales adyacentes. También la zona de servicio secundaria está libre de interferencia en el mismo canal, pero no de la proveniente de canales adyacentes. La zona de servicio primaria se define por el nivel de la intensidad del campo eléctrico de la onda terrestre. Este nivel varía de 0.1 mV/m en zonas rurales a 50 mV/m en una ciudad. Se considera servicio secundario el que se proporciona en zonas en las cuales la onda celeste (indirecta) excede 0.5 mV/m el 50% del tiempo o más. Una estación de canal libre clasificada como I-B tiene una potencia de salida del transmisor que varía de 10 a 50 kW, y más de una estación en el interior de Estados Unidos puede operar en la misma frecuencia, sujeta a las limitaciones de canal común y canales adyacentes. Para las estaciones clase I-A, se permite que por la noche funcione una sola estación dentro de un radio de 750 millas; de esta forma se evita la interferencia de canal común.

Las estaciones de canal libre también pueden clasificarse como de clase II, en cuyo caso la potencia de salida puede ser tan baja como 250 W, dependiendo de la subclasificación particular (es decir, clase II-A, II-B o II-C). Las estaciones clase II tienen una potencia de salida máxima de 50 kW, se consideran secundarias de las estaciones de clase I, y se diseñan para atender a una zona de servicio primaria sometida a y limitada por interferencias de las estaciones de clase I. Siempre que sea necesario, a fin de reducir la interferencia se utiliza una antena direccional, aunque también es posible reducir la potencia durante la noche.

De los 107 canales de radiodifusión estándar, 41 están diseñados para dar servicio como canales regionales. Varias estaciones pueden operar con potencia de salida del transmisor de 5 kW o menos en la misma frecuencia, sujetos a las limitaciones impuestas por el contorno de la intensidad de campo y la consecuencia de la interferencia. Una estación de clase III funciona a fin de prestar servicio a un centro de población y a la zona rural contigua a él. En la clase III existen dos subclasificaciones: la clase III-A, en la cual la potencia de salida varía de 1 a 5 kW inclusive, y la clase III-B, en la cual la potencia nocturna varía de 0.5 a 1 kW inclusive y la potencia máxima durante el día es de 5 kW.

El tercer tipo de canales, los locales, es para las estaciones de clase IV, diseñadas para prestar servicio a una ciudad o pueblo y a las zonas suburbana y rural contiguas a ellos. Tales estaciones operan con una potencia máxima de 1 kW durante el día y de 250 W por la noche, excepto las estaciones de Florida que se encuentran al sur de los 28° de latitud norte y entre los

meridianos de 80° y 82° de longitud oeste, en cuyo caso la potencia está limitada a 250 W durante el día y la noche. Algunas estaciones de esta clase ya estaban autorizadas para operar con una potencia de salida de 100 W.

#### Intensidad de campo

La intensidad de campo producida por una estación de radiodifusión estándar es un factor clave para determinar las zonas de servicio primaria y secundaria, así como las limitaciones debidas a interferencia de posibles estaciones de radio en el futuro. Las limitaciones de intensidad de campo son especificadas por la FCC con unidades de volts por metro; sin embargo, los dispositivos medidores pueden leer volts o decibeles con respecto a 1 mw (dBm), y puede ser necesaria una conversión a fin de obtener la intensidad de campo. La potencia recibida puede medirse en dBm y convertirse a watts. Las lecturas de voltaje pueden convertirse a watts elevando al cuadrado el voltaje eficaz o raíz cuadrático medio (rms, de *root mean square*) y dividiendo entre la resistencia de entrada del medidor de intensidad de campo, que por lo común es del orden de 50 o 75Ω. Otros factores necesarios para determinar la intensidad del campo eléctrico son la ganancia y las pérdidas de potencia del sistema de la antena receptora de la intensidad de campo. Una vez que se han determinado estos factores, el área efectiva con compensación de pérdidas en la antena receptora de la intensidad de campo puede calcularse mediante:

$$A_{ef} = G \frac{\lambda^2}{4\pi} L$$

donde  $A_{ef}$  = área efectiva incluyendo la compensación de pérdidas,  $m^2$

$G$  = ganancia de potencia en la antena receptora de la intensidad de campo, W/W

$\lambda$  = longitud de onda, m

$L$  = factor de pérdida por mal acoplamiento y pérdidas debidas al cableado, W/W

A partir de este cálculo es posible obtener la densidad de potencia por metro cuadrado dividiendo la potencia recibida entre el área efectiva, y la intensidad del campo eléctrico puede calcularse como:

$$E = \sqrt{\mathcal{P} \times Z_{el}}$$

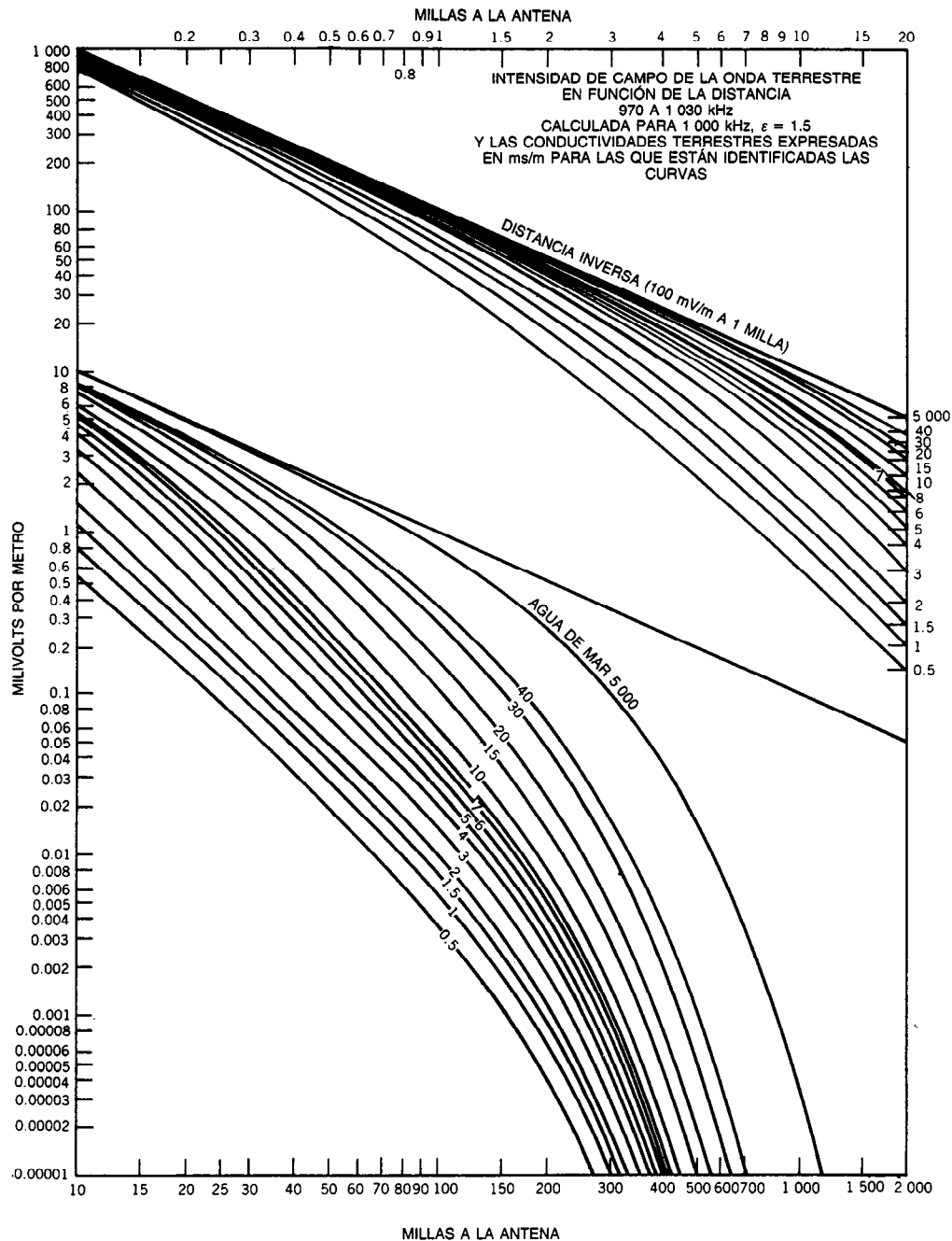
donde  $E$  = intensidad del campo eléctrico, V/m

$\mathcal{P}$  = densidad de potencia, W/m<sup>2</sup>

$Z_{el}$  =  $120\pi\Omega$ , impedancia del espacio libre

Las intensidades de campo necesarias para proporcionar servicio primario a zonas fabriles y de comercios en una ciudad varían de 10 a 50 mV/m; para zonas residenciales en una ciudad, las intensidades varían de 0.1 a 0.5 mV/m, excepto en las zonas del sur de Estados Unidos durante el verano, cuando el intervalo es de 0.25 a 1 mV/m.

En la tabla 49-5 se presentan los contornos de servicio protegido y los de la interferencia permisible para



**Fig. 49-14.** Curvas de propagación de las ondas terrestres. Tomado de la FCC, *Rules and Regulations*, Vol. II, parte 73, pág. 74.

estaciones de radiodifusión estándar. Las intensidades de campo que se proporcionan en esta tabla, junto con un conocimiento de las estaciones de radiodifusión existentes, pueden utilizarse para determinar el potencial de establecer nuevas estaciones de radiodifusión estándar.

#### Características de propagación

Uno de los factores de mayor importancia en la determinación de la intensidad de campo es la característica de propagación, descrita por el cambio en la intensidad de campo eléctrico con un incremento en la distancia a la antena de la estación radiodifusora. Esta

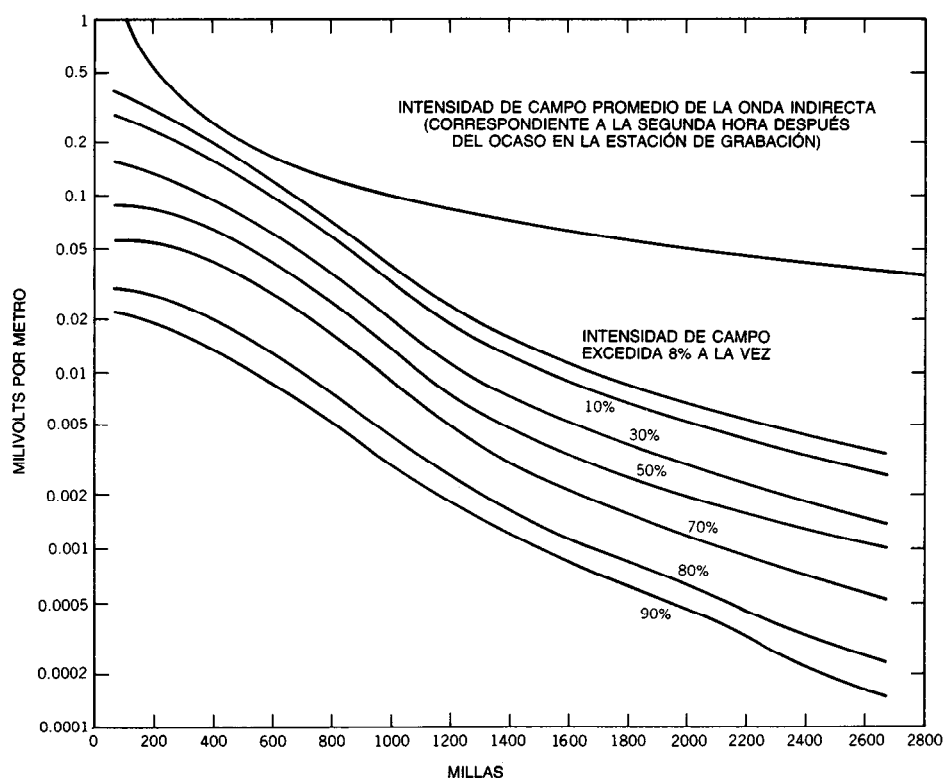


Fig. 49-15. Curvas de propagación de las ondas indirectas. Las curvas no se consideran suficientemente exactas para su utilización práctica en distancias menores de aproximadamente 250 millas. Tomado de la FCC, *Rules and Regulations*, Vol. III, parte 73, pág. 90.

variación depende de varios factores, incluyendo frecuencia, constante dieléctrica de la superficie, tangente de pérdida superficial, polarización, topografía local y hora del día. En términos generales, la propagación permite mayores intervalos y ocurre durante periodos nocturnos, por lo que es necesario que algunas estaciones reduzcan la potencia o no funcionen por la noche a fin de evitar la interferencia. Las curvas de propagación en el campo de radiodifusión suelen estar asociadas con un nivel de referencia de 100 mV/m en 1 milla; sin embargo, es posible obtener una expresión más general de la propagación de las ondas terrestres utilizando la serie de Bremmer. En la figura 49-14 se muestran curvas típicas de propagación de las ondas terrestres para la banda de radiodifusión estándar.

La potencia radiada efectiva (ERP, de *effective radiated power*) se refiere a la salida de potencia efectiva de la antena en una dirección específica, e incluye la salida de potencia del transmisor, pérdidas en la línea de transmisión y ganancia de potencia de la antena. La ERP suele exceder en la mayor parte de los casos la potencia de salida del transmisor. Para un supuesto radiador isotrópico perfecto, se encuentra que la ERP es:

$$ERP = \frac{E^2 r^2}{30}$$

donde  $E$  representa la intensidad de campo eléctrico en volts por metro y  $r$  es la distancia en metros.

Para una distancia de 1 milla (1 609.34 m), la ERP necesaria para producir una intensidad de campo de 100 mV/m es 863.3 W. Ya que la intensidad de campo es proporcional a la raíz cuadrada de la potencia, es posible determinar intensidades de campo a otras potencias.

La propagación de ondas celestes (o indirectas) implica necesariamente algo de desvanecimiento e intensidades de campo menos predecibles, y se describe de forma más idónea en términos de estadísticas o del porcentaje de tiempo que se encuentra una intensidad de campo específica. En la figura 49-15 se muestra la propagación de ondas indirectas para 1 MHz a latitudes de 36° a 50°.

#### Transmisores

Las reglas que rigen a los transmisores de radiodifusión AM se encuentran en la norma TR-101A de la Electronic Industry Association (EIA), *Electrical Performance Standard for Standard Broadcast Transmitters*. Los parámetros y métodos de medición abarcan lo siguiente: clasificación de salida de la portadora, capacidad de salida de potencia de la portadora, intervalo de frecuencia de la portadora, estabilidad de frecuencia de la portadora, desplazamiento de la portadora, nivel de



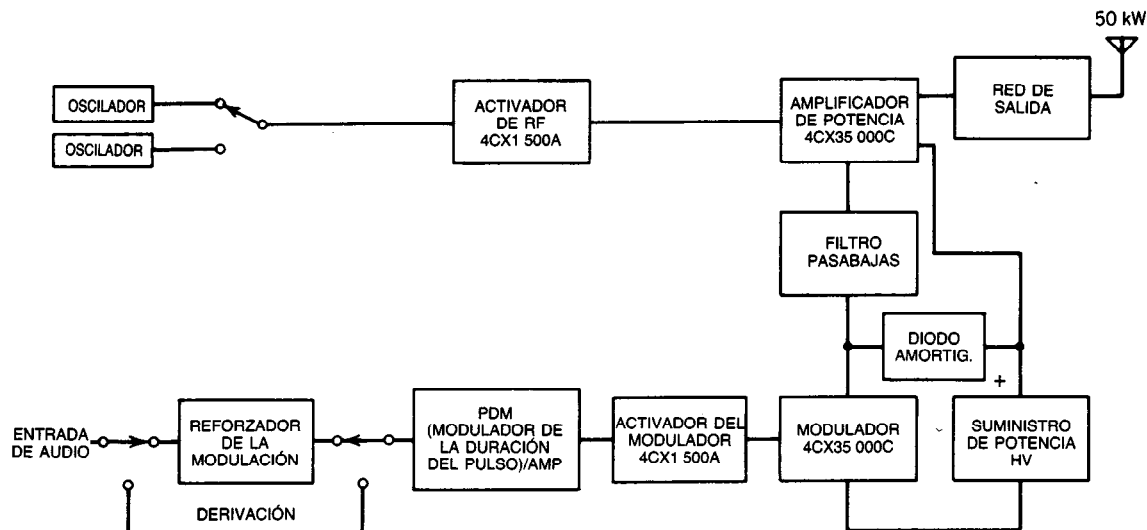


Fig. 49-16. Diagrama de bloques de un transmisor AM de 50 kW. Cortesía de la Harris Corporation.

ruido de la portadora, magnitud de los armónicos de radiofrecuencia, carga normal, facilidades para el ajuste del circuito de salida del transmisor, definiciones de las interfaces RF y de audio, capacidad de modulación, nivel de entrada de audio para el 100% de modulación, respuesta de la frecuencia de audio, distorsión armónica de la frecuencia de audio, alimentación nominal y su variación y entrada de potencia.

Los transmisores para radiodifusión AM estándar varían en cuanto a salida de potencia desde los de tipo de estado sólido de 250 W hasta las unidades de tubos al vacío de 50 kW. En la figura 49-16 se muestra un diagrama de bloques de un transmisor típico de radiodifusión de 50 kW.

#### Sistemas de antena

Los sistemas de antena para la radiodifusión AM estándar suelen constar de un grupo de elementos verticales de un cuarto de onda combinados en una red en fase con objeto de producir contornos de intensidad de campo eléctrico deseados. Por ejemplo, si un transmisor y un sistema de antena se colocan al norte de la población a la que prestan servicio, la energía radiada puede redistribuirse para obtener mejor cobertura al sur del transmisor. El sistema de antena debe estar bien acoplado con el transmisor con una onda estacionaria de voltaje menor que 1.2:1, a fin de evitar que se refleje una potencia excesiva de regreso al transmisor. Las ganancias típicas para los sistemas de antena de radiodifusión varían de 3 a 8 dB en una isotrópica perfectamente lineal. Las alturas típicas de cada vertical varían de 200 a 500 pies.

#### 49.2.2 Modulación de frecuencias (FM)

El término radiodifusión por modulación de frecuencia (FM, de *frequency modulation*) se refiere a la transmisión de habla y música, que el público general recibe

en la banda de frecuencias de 88 a 108 MHz. La modulación de frecuencia se utiliza para obtener mayor fidelidad de recepción que la existente con la radiodifusión estándar AM. En 1961 se introdujo la radiodifusión estereofónica, con la adición de una portadora suprimida de doble banda lateral para la transmisión de una señal de diferencia izquierda menos derecha. El canal de suma izquierda más derecha se envía empleando la modulación de frecuencia normal. Algunos sistemas de radiodifusión por FM también incluyen una subportadora de autorización de comunicación subsidiaria (SCA, de *subsidiary communication authorization*) para usos comerciales privados. La radiodifusión por modulación de frecuencia está limitada típicamente a transmisiones en línea rectal (visual). Como resultado, la cobertura por FM se localiza en un intervalo de aproximadamente 75 millas, dependiendo de la altura de la antena y de la potencia radiada efectiva.

#### Asignaciones de frecuencias

Las 100 frecuencias portadoras para la radiodifusión por FM varían de 88.1 a 107.9 MHz, y son equidistantes cada 200 kHz. Los canales de 88.1 a 91.9 MHz se utilizan para la radiodifusión no comercial, y los que están de 92.1 a 107.9 MHz, para la radiodifusión comercial. Cada canal tiene un ancho de banda de 200 kHz, y se incluye una banda de guarda o de seguridad de 25 kHz que permite una desviación de frecuencia máxima de  $\pm 75$  kHz. La frecuencia portadora debe mantenerse con  $\pm 2\,000$  Hz.

#### Clasificaciones de las estaciones

En la radiodifusión por FM, las estaciones se clasifican como de clase A, clase B o clase C. Las estaciones de clase A prestan servicio a ciudades o pueblos relativamente pequeños y sus alrededores, y están limitadas a una potencia radiada efectiva máxima de 3 kW.

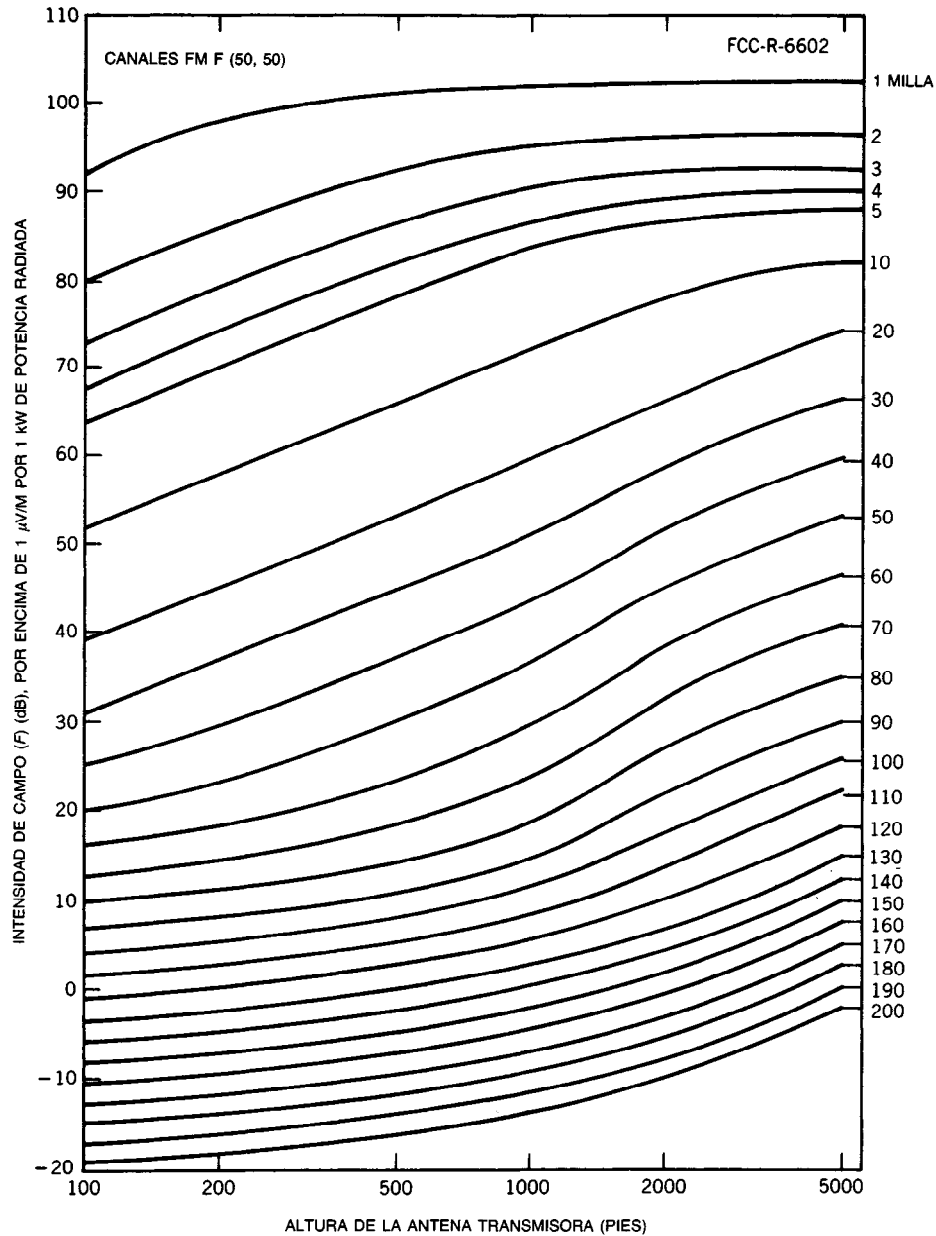
Tabla 49-5. Intensidades para la señal de servicio protegido e intensidades permisibles de interferencia para radiodifusión estándar (AM)

Clase de estación	Clase de canal utilizado	Potencia permisible	Contorno de intensidad de señal del área protegida contra la interferencia perjudicial <sup>tr</sup>		Señal de interferencia permisible en el mismo canal	
			Día <sup>b</sup>	Noche	Día <sup>b</sup>	Noche <sup>c</sup>
I-A	Libre	50 kW	$\left\{ \begin{array}{l} \text{SC } 100 \text{ uV/m} \\ \text{AC } 500 \text{ uV/m} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{SC } 500 \text{ uV/m(50\% onda indirecta)}^d \\ \text{AC } 500 \text{ uV/m}^b \end{array} \right\}$	5 uV/m	25 uV/m <sup>d</sup>
I-B	Libre	10 a 50 kW	$\left\{ \begin{array}{l} \text{SC } 100 \text{ uV/m} \\ \text{AC } 500 \text{ uV/m} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{SC } 500 \text{ uV/m } 50\% \text{ onda indirecta} \\ \text{AC } 500 \text{ uV/m}^b \end{array} \right\}$	5 uV/m	25 uV/m
II-A	Libre	$\left\{ \begin{array}{l} 0.25 \text{ a } 50 \text{ kW (de día)} \\ 10 \text{ a } 50 \text{ kW (de noche)} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{AC } 500 \text{ uV/m} \\ 500 \text{ uV/m} \end{array} \right\}$	500 uV/m <sup>b</sup>	25 uV/m	25 uV/m
II-B y II-D	Libre	0.25 a 50 kW	500 uV/m	2 500 uV/m <sup>b,e</sup>	25 uV/m	125 uV/m
II-B y II-D <sup>f</sup>	Libre	0.25 a 1 kW <sup>f</sup>	500 uV/m	10 000 uV/m <sup>f</sup>	25 uV/m	500 uV/m <sup>f</sup>
III-A	Regional	1 a 5 kW	500 uV/m	2 500 uV/m <sup>b</sup>	25 uV/m	25 uV/m
III-B	Regional	0.5 a 1 kW (noche) y 5 kW (día)	500 uV/m	4 000 uV/m <sup>b</sup>	25 uV/m	200 uV/m
IV	Local	0.25 kW (noche) y 0.25 a 1 kW (día)	500 uV/m	Sin prescribir	25 uV/m	Sin prescribir

Fuente: Parte 73.183 de *Rules and Regulations*, de la Federal Communications Commission.

SC, mismo canal; AC, canal adyacente.

<sup>a</sup> Cuando una estación ya está limitada por la interferencia de otras estaciones a un contorno de mayor valor que el normalmente protegido para su clase, este contorno debe establecerse estándar para tal estación con respecto a la interferencia de todas las demás estaciones.<sup>b</sup> Onda terrestre.<sup>c</sup> Intensidad de campo de la onda indirecta durante el 10% del tiempo o más.<sup>d</sup> Las estaciones clase I-A en canales reservados para uso exclusivo de una estación durante las horas nocturnas están protegidas contra la interferencia de canal común con este fundamento.<sup>e</sup> Los valores son con respecto a la interferencia de todas las estaciones, excepto las de clase I-B, que pueden provocar interferencia a un contorno de intensidad de campo de valor superior. Sin embargo, se recomienda que las estaciones de clase II sean asignadas de modo que la interferencia recibida de las estaciones de clase I-B no exceda estos valores. Si las estaciones de clase II están limitadas por las estaciones de clase I-B a valores superiores, entonces tales valores deben establecerse estándar con respecto a la protección contra todas las demás estaciones.<sup>f</sup> Se aplica solamente a operaciones nocturnas de las estaciones clase II-B que vienen en §73.21(a)(2)(ii)(C), y al funcionamiento de las estaciones de clase II-D limitadas en tiempo durante horas de la noche diferentes de las autorizadas para funcionar el 1.º de junio de 1980.



**Fig. 49-17.** Intensidad de campo de FM estimada para una altura de antena receptora de 30 pies.  $F(50, 50)$ , intensidad de campo para el 50% del tiempo en el 50% de localizaciones. Tomado de la FCC, *Rules and Regulations*, Vol. III, parte 73, pág. 150.

Asimismo, las estaciones de clase A están limitadas a 20 frecuencias entre 92.1 y 107.1 MHz inclusive. Las estaciones de clase B y C prestan servicio a ciudades o pueblos más grandes y sus alrededores. A las estaciones de clase B se les permite operar a una potencia radiada efectiva máxima de 50 kW con una antena cuya altura efectiva sea de 500 pies. Para alturas mayores de 500 pies por encima del terreno en el que se encuentran enclavadas las antenas, se debe reducir la

potencia. Las estaciones de clase C operan a una potencia radiada efectiva de 100 kW y con antenas hasta de 2 000 pies de altura sin reducciones de potencia.

#### Intensidad de campo y propagación

La intensidad de campo producida por una estación de radiodifusión por FM depende de la potencia radiada efectiva, altura de la antena, terreno local, condiciones de dispersión troposférica y otros factores. Sin embar-

Tabla 49-6. Guías para la separación mínima de las estaciones de FM

Clase de estación	Separación en millas		Instalaciones autorizadas	
	Canal común	Primera adyacente	Potencia (kW)	Altura de la antena (pies)
A a A	45 a 65		3	300
A a A	40 a 44		2	300
A a A	Menos de 40		1	300
A a B		50 a 65	3	Clase A 300
			50	Clase B 500
A a B		40 a 49	3	Clase A 300
			20	Clase B 500
A a B		Menos de 40	3	Clase A 300
			10	Clase B 500
A a C		80 a 105	3	Clase A 300
			100	Clase C 2 000
A a C		60 a 79	3	Clase A 300
			50	Clase C 2 000
A a C		Menos de 60	3	Clase A 300
			20	Clase C 2 000
B a B	125 a 150	80 a 105	50	500
B a B	100 a 124	65 a 79	20	500
B a B	75 a 99	50 a 64	10	500
B a B	Menos de 75	Menos de 50	5	500
B a C	140 a 170	110 a 135	50	Clase B 500
			100	Clase C 2 000
B a C	110 a 139	85 a 109	20	Clase B 500
			50	Clase C 2 000
B a C	90 a 100	60 a 84	10	Clase B 500
			20	Clase C 2 000
B a C	Menos de 90	Menos de 60	5	Clase B 500
			10	Clase C 2 000
C a C	150 a 180	125 a 150	100	2 000
C a C	120 a 149	95 a 124	50	2 000
C a C	100 a 119	75 a 94	20	2 000
C a C	Menos de 100	Menos de 75	10	2 000

Fuente: Parte 73, págs. 128 y 129 de *Rules and Regulations* de la Federal Communications Commission.

go, desde un punto de vista estadístico, es posible obtener una estimación de la intensidad de campo (fig. 49-17). Un factor para determinar las nuevas autorizaciones para la radiodifusión por FM es la separación en millas de las estaciones de canal común y de canal adyacente, la clase de estación y la altura de la antena. En la tabla 49-6 se presentan orientaciones típicas. Además, los contornos de la intensidad de campo se estiman para niveles de 3.16 mV/m (70 dBμ) y 1 mV/m (60 dBμ). Éstos pueden determinarse a partir de gráficas parecidas a la que se muestra en la figura 49-17. La propagación a frecuencias de FM (88-108 MHz) se concibe en la generalidad de los casos como una línea recta; sin embargo, unas mayores potencias radiadas efectivas, junto con los efectos de difracción, refracción y dispersión troposférica, permiten una cobertura ligeramente más amplia que la línea recta.

#### Transmisores

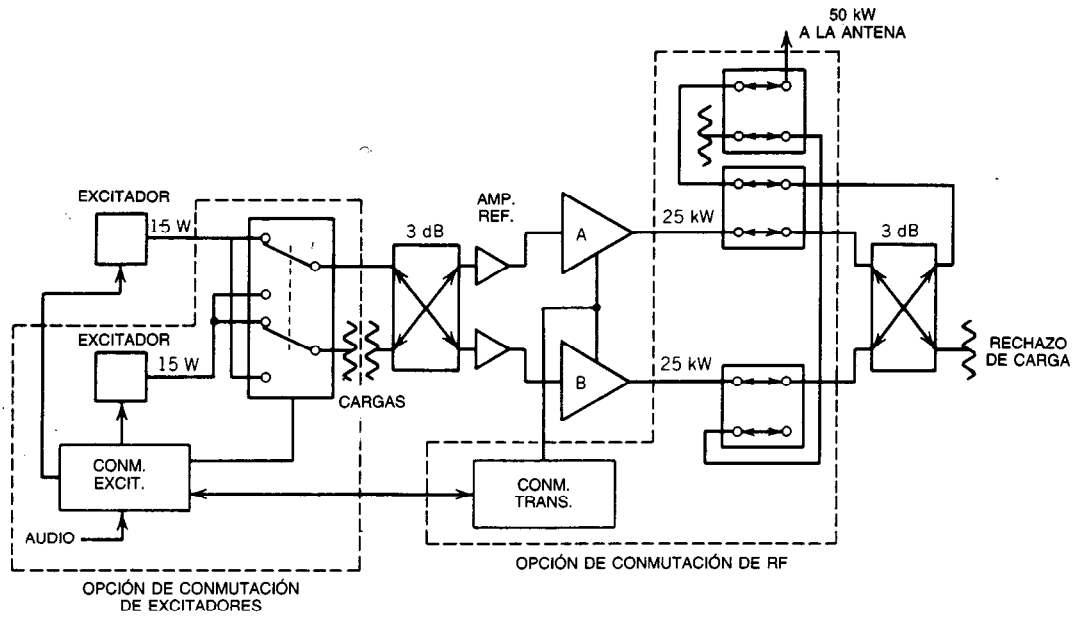
Los transmisores para la radiodifusión por FM suelen variar en cuanto a potencia de salida desde los de esta-

do sólido de 300 W hasta los de tipo tubo de vacío de 100 kW. En la figura 49-18 se muestra un diagrama de bloques de un transmisor dual por FM. Este sistema consta de dos transmisores de 25 kW que se hacen funcionar en paralelo y que proporcionan mayor confiabilidad en caso de alguna falla en el excitador o en el amplificador del transmisor.

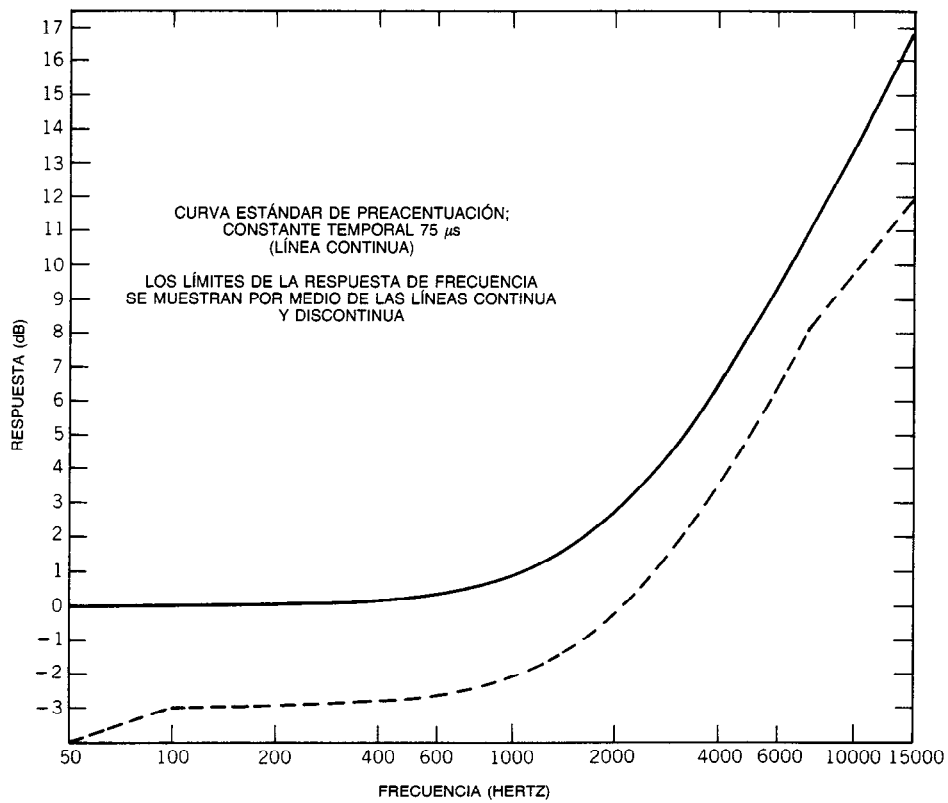
#### Sistemas de antena

Los sistemas de antena para la radiodifusión por FM deben tener una componente horizontal de polarización. Sin embargo, la mayor parte de los sistemas de antena están polarizados circularmente, por lo que cuentan con componentes tanto vertical como horizontal.

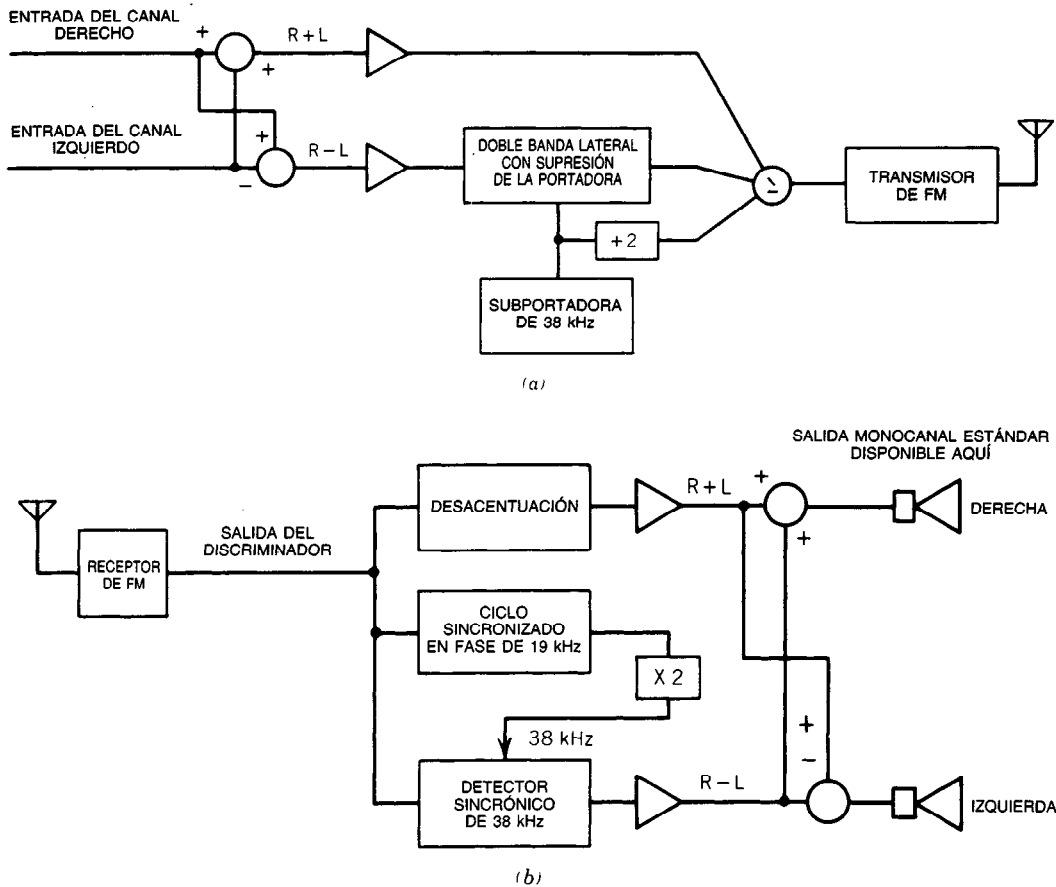
Típicamente el sistema de antena concentra la energía en el plano horizontal, con lo que se obtiene una mayor potencia radiada efectiva que resulta de las ganancias de potencia hasta 10 dB aproximadamente. Esto significa que un transmisor de 5 kW podría tener una potencia radiada efectiva de 50 kW. Además, po-



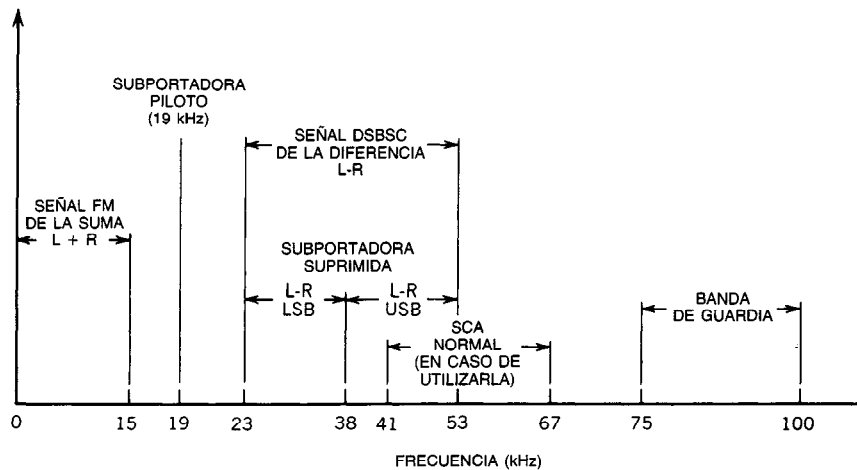
**Fig. 49-18.** Diagrama de bloques de un transmisor dual de FM. Cortesía de la Harris Corporation. AMP. REF. = amplificadores de refuerzo; CONM. EXCIT. = conmutador de los excitadores. CONM. TRANS. = conmutador de los transmisores.



**Fig. 49-19.** Límites de preacentuación basados en una constante de tiempo de  $75 \mu s$ . Tomado de la FCC, *Rules and Regulations*, Vol. III, parte 73, pág. 153.



**Fig. 49-20.** Diagrama de bloques de un sistema estereofónico de FM: a) transmisor; b) receptor. R = derecho; L = izquierdo.



**Fig. 49-21.** Localizaciones del espectro para radiodifusión por FM utilizando subportadoras multicanalizadas de comunicación subsidiaria (SCA). DSBSC = transmisión de doble banda lateral con supresión de portadora; L = izquierda; R = derecha; LSB = banda lateral inferior; USB = banda lateral superior.

dría permitirse un patrón direccional en el plano horizontal con hasta 15 dB de variación a fin de acentuar la cobertura en zonas predeterminadas.

### Preacentuación

La preacentuación o preénfasis se emplea en un transmisor de radiodifusión por FM a fin de mejorar la razón de señal sobre ruido. El límite superior de la preacentuación que se muestra en la figura 49-19 está basado en una constante temporal de 75  $\mu$ s, como requiere la FCC para los transmisores de radiodifusión por FM. Las audiofrecuencias de 50 a 2 120 Hz se transmiten esencialmente con la FM normal, mientras que las audiofrecuencias de 2 120 Hz a 15 kHz se acentúan con un mayor índice de modulación. Existe bastante mejora de la razón de señal sobre ruido cuando el receptor está equipado con un circuito de acoplamiento desacentuador. En la figura 49-19 también se muestra el límite inferior para la preacentuación.

### Espectro de FM

El sistema monofónico fue desarrollado inicialmente para permitir que las transmisiones de sonido para audiofrecuencias de 50 a 15 000 Hz estuviesen contenidas dentro de un ancho de banda de RF de  $\pm 75$  kHz. Con el desarrollo de la radiodifusión estereofónica por FM, la señal original FM (que consta de un canal derecho más izquierdo) se transmite en un ancho de banda más pequeño a fin de hacerla compatible con un receptor monofónico de FM, y un canal derecho menos izquierdo es multicanalizado en frecuencia en una subportadora de 38 kHz mediante el empleo de doble banda lateral con supresión de la portadora. De la subportadora de 38 kHz se deriva una subportadora no modulada de 19 kHz, a fin de obtener una referencia de demodulación sincrónica para el receptor estereofónico. En la figura 49-20 se muestra un diagrama de bloques funcional de un sistema estereofónico por FM. El detector sincrónico recupera en los 38 kHz la información del canal izquierdo menos derecho, que entonces se combina con la información del canal izquierdo más derecho en combinadores de suma y diferencia para producir las señales originales de los canales izquierdo y derecho. Además, se utiliza una subportadora multicanalizada de autorización de comunicación subsidiaria (SCA) con objeto de ofrecer música de fondo a los suscriptores comerciales. La subportadora SCA limita el intervalo entre 20 y 75 kHz; los 41 y 67 kHz son los más populares. La modulación SCA está limitada al 30% de la portadora principal para monofonía y al 10% para estereofonía. En la figura 49-21 se muestra el espectro para una radiodifusión por FM típica en la que se utilizan estereofonía y SCA.

#### 49.2.3 Emisiones televisivas

Las emisiones televisivas consisten en transmisiones de imagen y sonido para el público general en las bandas de frecuencias muy alta (VHF, de *very-high-frequency*) y ultraalta (UHF, de *ultrahigh-frequency*). En el sistema compuesto se utilizan portadoras separadas para la

imagen, el sonido y la información del color, diseñadas de forma que sea posible obtener recepción compatible en un receptor de blanco y negro o de color.

### Asignaciones de frecuencias

En Estados Unidos, las asignaciones de frecuencias se dividen en 68 canales en las bandas VHF y UHF. Para la VHF, los canales del 2 al 13 incluyen de 54 a 72, de 76 a 88, y de 174 a 216 MHz. Para la UHF, los canales del 14 al 69 abarcan de 470 a 806 MHz, y los canales del 70 al 83 incluyen de 806 a 890 MHz. Cada canal consta de una banda de frecuencia de 6 MHz que contiene modulación de imagen, de audio, de color y posiblemente una modulación especial utilizada por espectadores sordos. En la tabla 49-7 se presentan las designaciones del canal y de la banda de frecuencia. Algunos de los canales son compartidos con otros servicios, incluyendo la telemetría biomédica, radio móvil y radioastronomía.

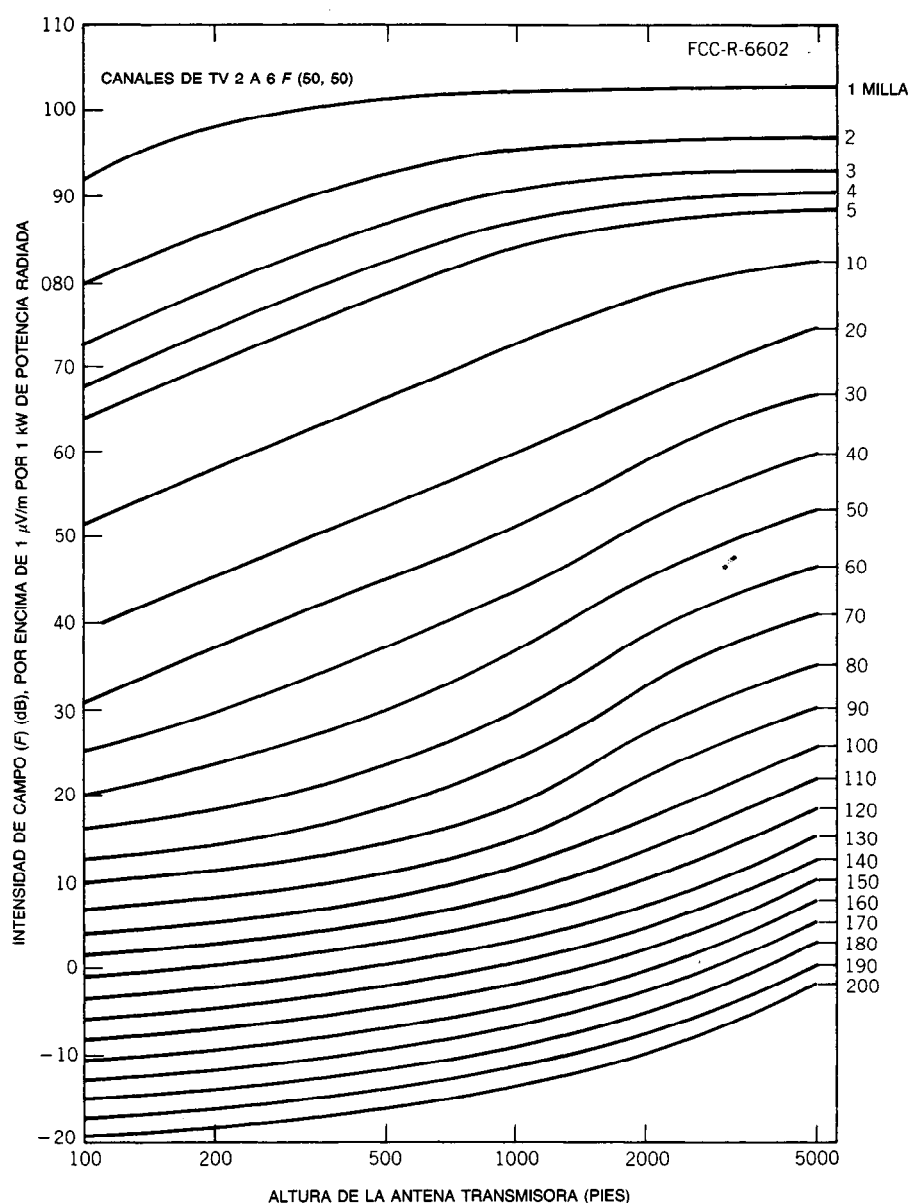
### Intensidad de campo y características de propagación

La intensidad de campo de un transmisor de emisiones televisivas depende de la distancia, altura de la antena, terreno local, frecuencia y potencia radiada efectiva. En la figura 49-22 se muestra una gráfica de la intensidad de campo de los canales 2 a 6. Las características de propagación de otros canales se proporcionan en la parte 73.699 del documento *Rules and Regulations* de la FCC. La autorización de nuevas estaciones depende de la separación mínima de asignaciones de canal común y varía de 170 a 220 millas para los canales del 2 al 13 y de 155 a 205 millas para los canales del 14 al 83, dependiendo de la localización. Para las asignaciones de canal adyacente, la distancia mínima es 60 millas para los canales del 2 al 13 y 55 millas para los canales del 14 al 88. Este requisito de separación mínima no siempre protege contra la interferencia provocada por la concesión de una nueva estación o el permiso para modificar una ya existente.

### Normas de transmisión

Se requiere que un transmisor de televisión tenga un mínimo de 100 W de potencia visual en el plano horizontal con polarización horizontal; la potencia máxima está determinada por los valores que se observan en la tabla 49-8. La portadora visual suele estar localizada 1.25 MHz por encima del límite inferior del canal. La subportadora del sonido se encuentra 4.5 MHz por encima de la portadora visual, y la subportadora de color está 63/88 partes de 5 MHz (3.57954545... MHz) por encima de la portadora visual, con una tolerancia de  $\pm 10$  Hz y una velocidad de deriva que no debe exceder 0.1 Hz/s. En la figura 49-23 se muestra un espectro típico.

El número de líneas de exploración por imagen (cuadro) es 525, entrelazadas dos a uno en campos sucesivos (262.5 líneas por campo) con una frecuencia de exploración horizontal de 15 750 Hz para monocromático y  $15\,734.264 \pm 0.044$  Hz para color. La frecuencia de exploración horizontal para color se obtiene dividiendo la frecuencia de la subportadora de sonido, 4.5 MHz, entre 286. La razón alto/ancho del cuadro es



**Fig. 49-22.** Intensidad de campo para los canales del 2 al 6 para una altura de antena receptora de 30 pies.  $F(50, 50)$ , intensidad de campo para el 50 % del tiempo en el 50 % de las localizaciones. Tomado de la FCC, *Rules and Regulations*, Vol. III, parte 73, pág. 226.

de 4 unidades horizontalmente a 3 unidades verticalmente. Las ondas de forma de sincronización para transmisiones a color están dadas en la figura 49-24, y para blanco y negro, en la figura 49-25.

La frecuencia de exploración vertical es 60 Hz para monocromático y 59.94 Hz para color. Esa última frecuencia puede obtenerse dividiendo la frecuencia de exploración horizontal de 15 734.264 entre el número de líneas por campo, 262.5. El nivel de borrado se transmite a  $75 \pm 2.5\%$  del nivel de cresta de la porta-

dora, y el nivel de referencia del negro se encuentra separado del nivel de borrado por  $7.5 \pm 2.5\%$  del intervalo de imagen (vídeo) desde el nivel de borrado hasta el nivel de referencia del blanco. Una disminución en la intensidad inicial de la luz provoca un aumento en la potencia radiada. La modulación de la imagen es una forma de modulación de la amplitud (AM) en la que la banda lateral inferior está significativamente atenuada y se denomina modulación residual. Esto permite la transmisión de la señal de



Tabla 40-7. Canales de televisión y bandas de frecuencia

Canal núm.	Banda de frecuencia (MHz)	Canal núm.	Banda de frecuencia (MHz)
2	54-60	43	644-650
3	60-66	44	650-656
4	66-72	45	656-662
5	76-82	46	662-668
6	82-88	47	668-674
7	174-180	48	674-680
8	180-186	49	680-686
9	186-192	50	686-692
10	192-198	51	692-698
11	198-204	52	698-704
12	204-210	53	704-710
13	210-216	54	710-716
14	470-476	55	716-722
15	476-482	56	722-728
16	482-488	57	728-734
17	488-494	58	734-740
18	494-500	59	740-746
19	500-506	60	746-752
20	506-512	61	752-758
21	512-518	62	758-764
22	518-524	63	764-770
23	524-530	64	770-776
24	530-536	65	776-782
25	536-542	66	782-788
26	542-548	67	788-794
27	548-554	68	794-800
28	554-560	69	800-806
29	560-566	70	806-812
30	566-572	71	812-818
31	572-578	72	818-824
32	578-584	73	824-830
33	584-590	74	830-836
34	590-596	75	836-842
35	596-602	76	842-848
36	602-608	77	848-854
37	608-614	78	854-860
38	614-620	79	860-866
39	620-626	80	866-872
40	626-632	81	872-878
41	632-638	82	878-884
42	638-644	83	884-890

Fuente: Parte 73, págs. 171 y 172 de *Rules and Regulations* de la Federal Communications Commission.

imagen dentro de aproximadamente la mitad del ancho de banda de RF requerido con la AM ordinaria. La atenuación de la banda lateral inferior se inicia a 0.75 MHz por debajo de la portadora de imagen; por consiguiente, la banda lateral inferior no está atenuada por completo.

El sonido se transmite en una subportadora que se encuentra 4.5 MHz por encima de la portadora de imagen. La modulación que se utiliza es FM con preénfasis a fin de mejorar la razón de señal sobre ruido recibida.

Para la transmisión en color, la información de la imagen se procesa a fin de obtener señales por separado para el contenido de la imagen sin color (luminancia) y para el contenido de color (crominancia). En la figura 49-26 se muestra un diagrama funcional del procesamiento de premodulación estándar de señales de televisión a color.

Una cámara de televisión en color produce tres señales simultáneas: una roja, otra azul y otra verde. Estas tres señales se combinan conforme a las normas de la industria ("NTSC Signal Specifications", 1954) a

Tabla 49-8. Potencias máximas para televisión

Canales núms.	Máxima potencia radiada efectiva visual (dB) por encima de un kW (dBk)
2 a 6	20 dBk (100 kW)
7 a 13	25 dBk (316 kW)
14 a 83	37 dBk (5 000 kW) <sup>a</sup>

Fuente: Parte 73, pág. 183 de *Rules and Regulations* de la Federal Communications Commission.

<sup>a</sup> La máxima potencia radiada efectiva visual de estaciones televisivas que operan en los canales 14 a 83 dentro de 250 millas de la frontera canadiense con Estados Unidos no puede exceder de 30 dBk (1 000 kW).

la referencia de fase se obtiene de la ráfaga de sincronización cromática, que se deriva de la subportadora de crominancia sin modulación y forma parte de la señal de sincronización que se muestra en la figura 49-24.

#### Transmisores

Un sistema de transmisión por televisión contiene dos transmisores: uno para la portadora de imagen y otro para la portadora de sonido. La amplitud de la portadora de la imagen se modula con una señal de vídeo compuesta (señal de imagen combinada con sincronización y subportadora de color). También se modula la frecuencia de la portadora del transmisor. Después de la modulación y amplificación, estas dos portadoras

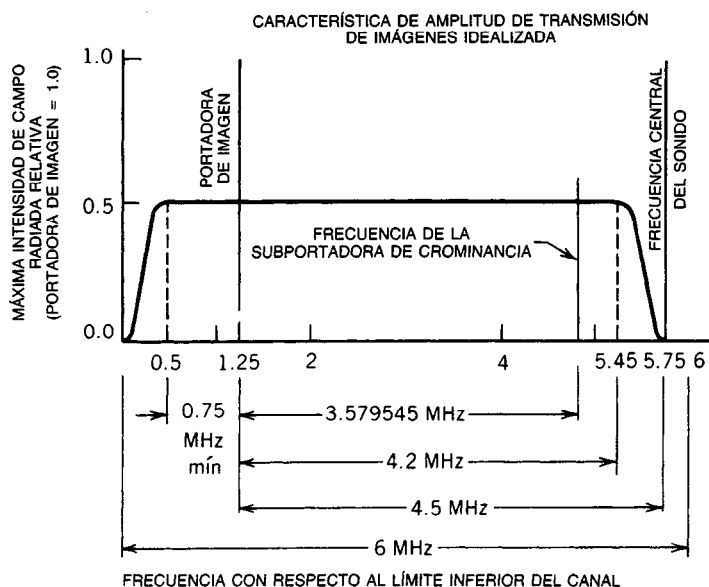


Fig. 49-23. Espectro de televisión (no está trazado a escala).

fin de producir una señal de luminancia (designada señal  $Y$ ) y dos señales de crominancia ( $I$  y  $Q$ ). Las señales  $I$  y  $Q$  modulan dos componentes de la subportadora de crominancia de 3.58 MHz que se encuentran en cuadratura de fase. Se utiliza la modulación de doble banda lateral con supresión de la portadora (DSBSC, de *double sideband suppressed-carrier*) (sec. 49.1), y las salidas de los dos moduladores se suman a la señal de luminancia.

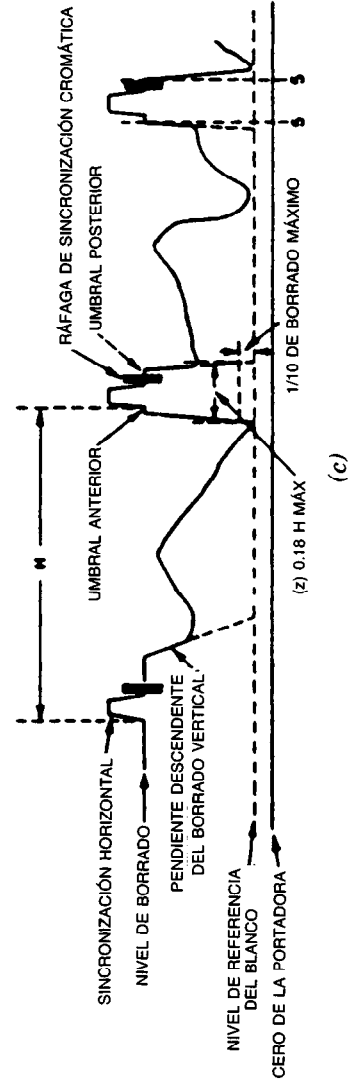
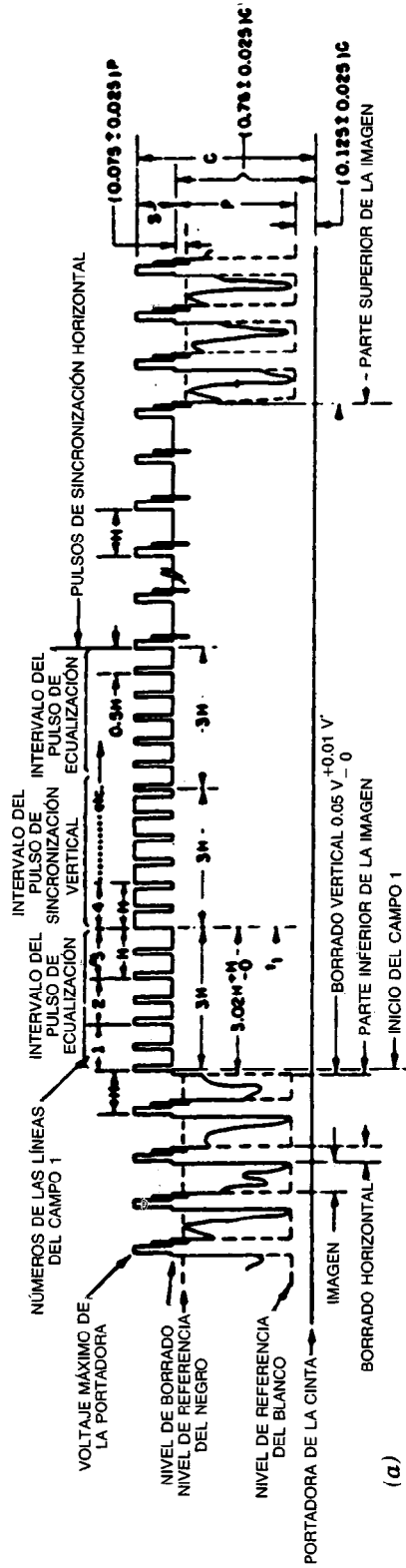
En una señal de imagen compuesta, que modula la portadora de imagen, la señal de luminancia se aplica directamente al modulador. Esta componente de la señal es compatible con los receptores de televisión en blanco y negro. La fase instantánea de la señal de la subportadora de crominancia es función del color de la imagen en ese instante, y la intensidad (saturación) del color está indicada por la amplitud instantánea.

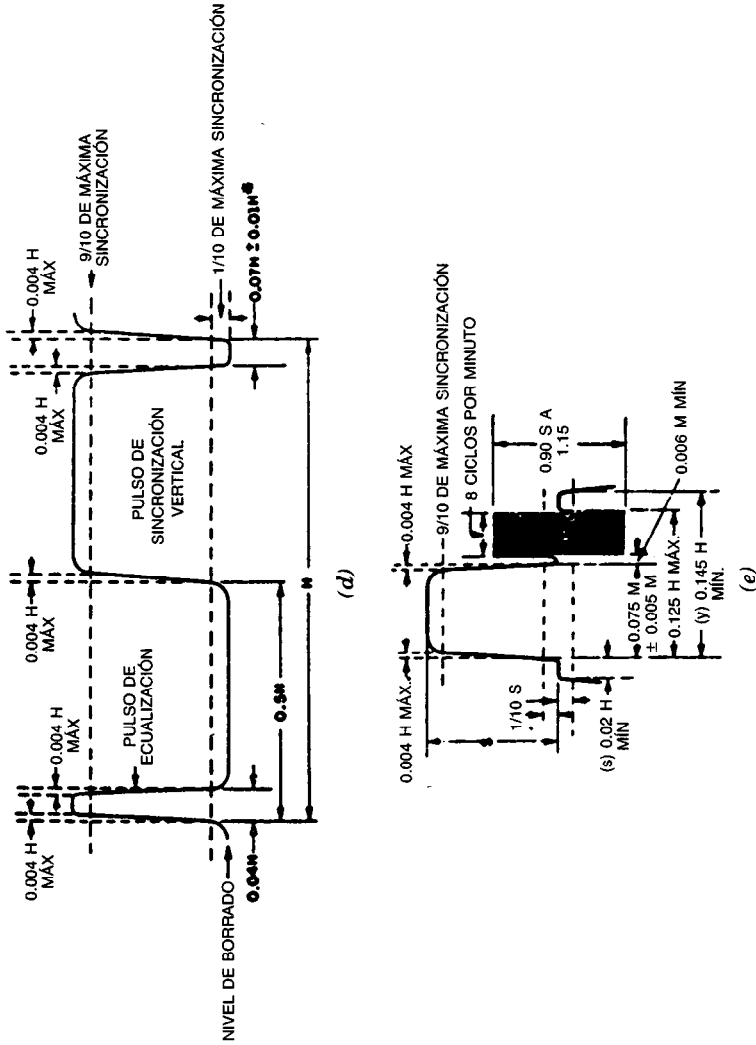
La desmodulación de la subportadora de crominancia en el receptor requiere una referencia de fase sincronizada y un detector sincrónico. La información de

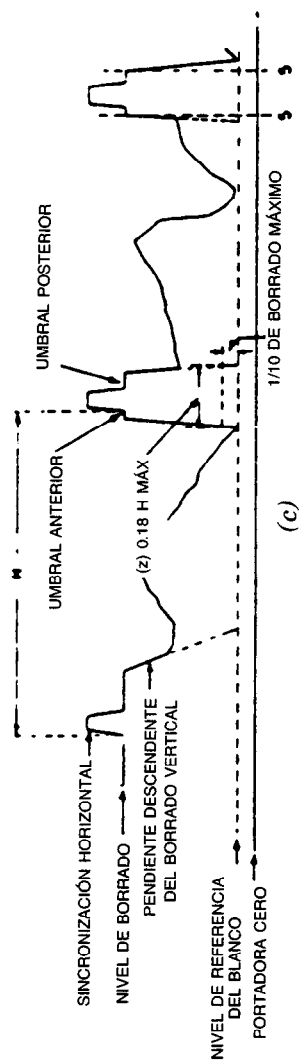
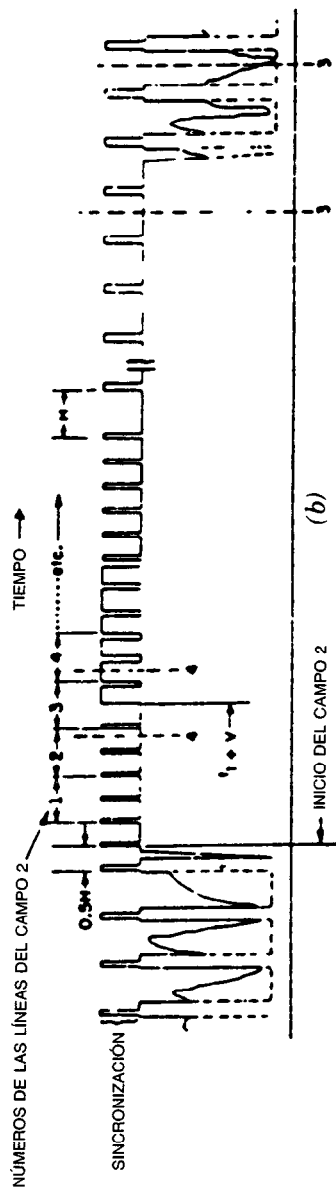
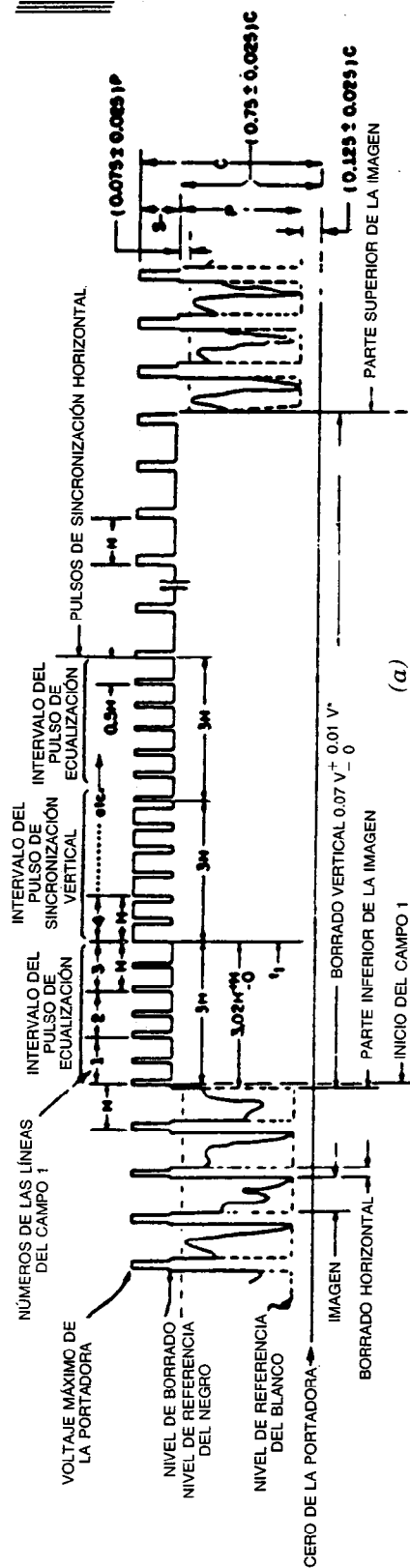
se combinan en una señal de televisión cuyo espectro se muestra en la figura 49-23. En la figura 49-27 se observa un diagrama de bloques funcional de un transmisor de televisión.

Los filtros deben insertarse después de los amplificadores de potencia y en los combinadores a fin de prevenir las interferencias a la señal de sonido por la señal de imagen y viceversa. Además, los filtros tienen la capacidad de suprimir los armónicos de las portadoras de radiofrecuencia.

El filtrado de la banda lateral residual puede efectuarse a bajos niveles de potencia si se utiliza un amplificador de potencia lineal para la portadora de imagen. Si se utiliza un amplificador RF clase C para alcanzar mayor eficiencia en el amplificador final de potencia, el filtro de banda lateral residual debe colocarse después de éste. El último esquema proporciona la máxima eficiencia de potencia en el amplificador final, pero requiere que el filtro de banda lateral residual disipe mayores cantidades de potencia.







**Fig. 49-25.** Ondas de forma para la sincronización en blanco y negro: *a*) campo 1; *b*) campo 2; *c*) detalle entre 3 y 4 en (*b*); *d*) detalle entre 4 y 5 en (*b*); *e*) detalle entre 5 y 6 en (*c*). En los incisos (*a*), (*b*) y (*c*) no están a escala las dimensiones horizontales. *H* = duración desde el inicio de una línea hasta el inicio de otra; *V* = duración desde el inicio de un campo hasta el inicio del siguiente; *P* = desplazamiento máximo de la señal de luminancia desde el nivel de borrado; *S* = amplitud de la sincronización por encima del nivel de borrado. El asterisco indica que las tolerancias dadas solamente se permiten para grandes variaciones temporales y no para ciclos consecutivos. Los flancos ascendente y descendente del borrado vertical deben completarse en menos de 0.1 *H*. Las pendientes ascendente y descendente del borrado horizontal deben tener el suficiente incremento (escalón) a fin de preservar los valores mínimo y máximo de (*x* + *y*) y (*z*) en todas las condiciones de contenido de la imagen. El área del pulso de equalización está entre 0.45 y 0.5 del área de un pulso de sincronización horizontal.

El inicio del campo 1 se define por medio de una línea completa entre el primer pulso de equalización y los pulsos de sincronización  $H$  precedentes. El inicio del campo 2 se define mediante la mitad de una línea entre el primer pulso de equalización y los pulsos de sincronización  $H$  precedentes. Los números de las líneas del campo 1 se inician con el primer pulso de equalización en el campo 1; los números de las líneas del campo 2 se inician con el segundo pulso de equalización en el campo 2.

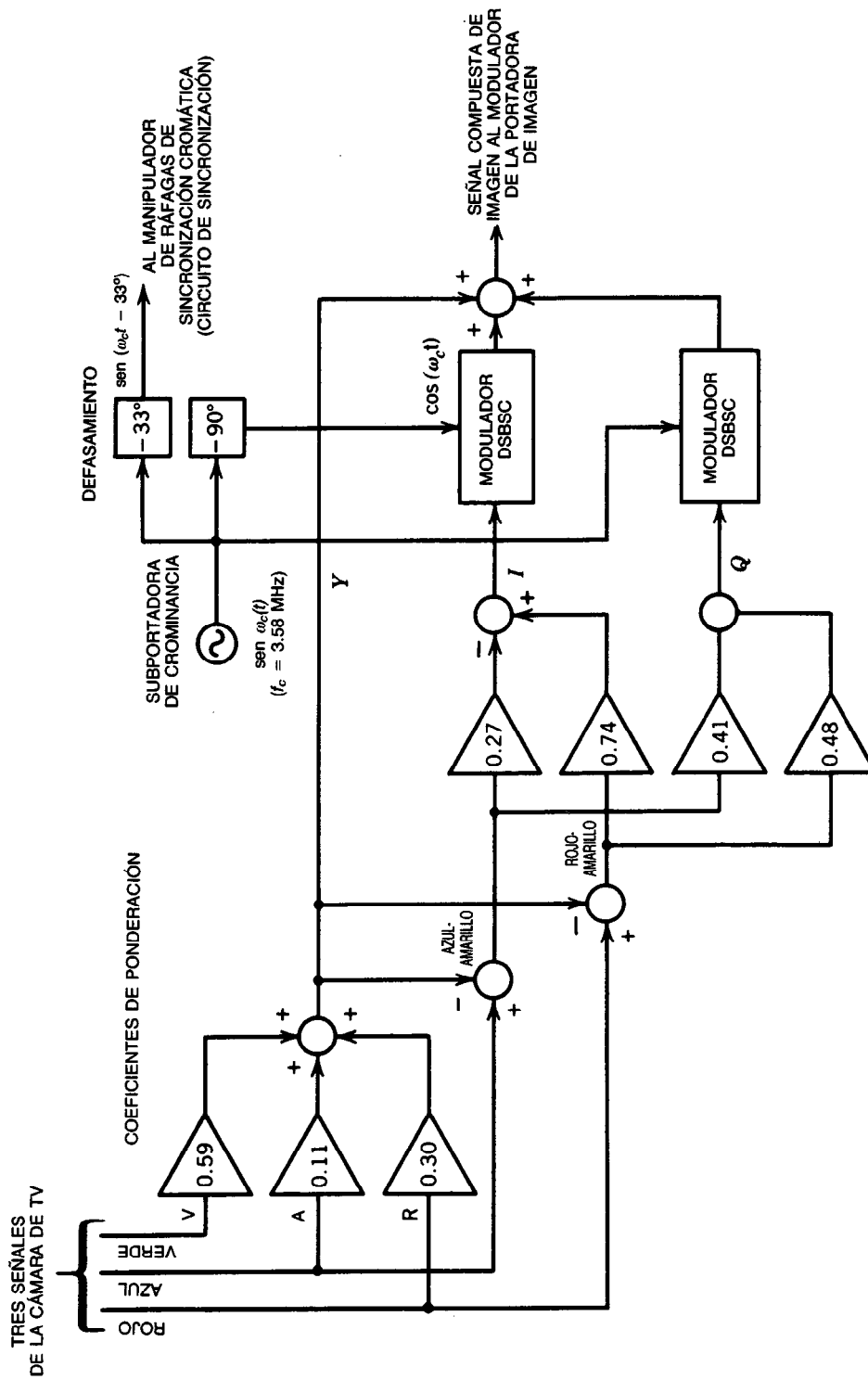


Fig. 49-26. Procesamiento de premodulación estándar para las señales de televisión en color. DSBSC = doble banda lateral con supresión de la portadora;  $Y$  = señal de luminancia;  $I$  y  $Q$  = señales de crominancia.

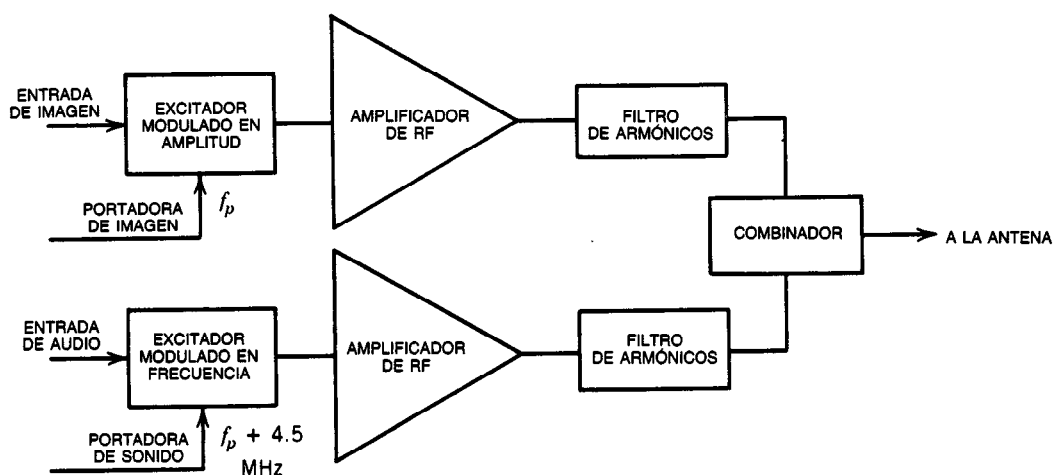


Fig. 49-27. Diagrama de bloques funcional de un transmisor de televisión. Cortesía de la Harris Corporation.

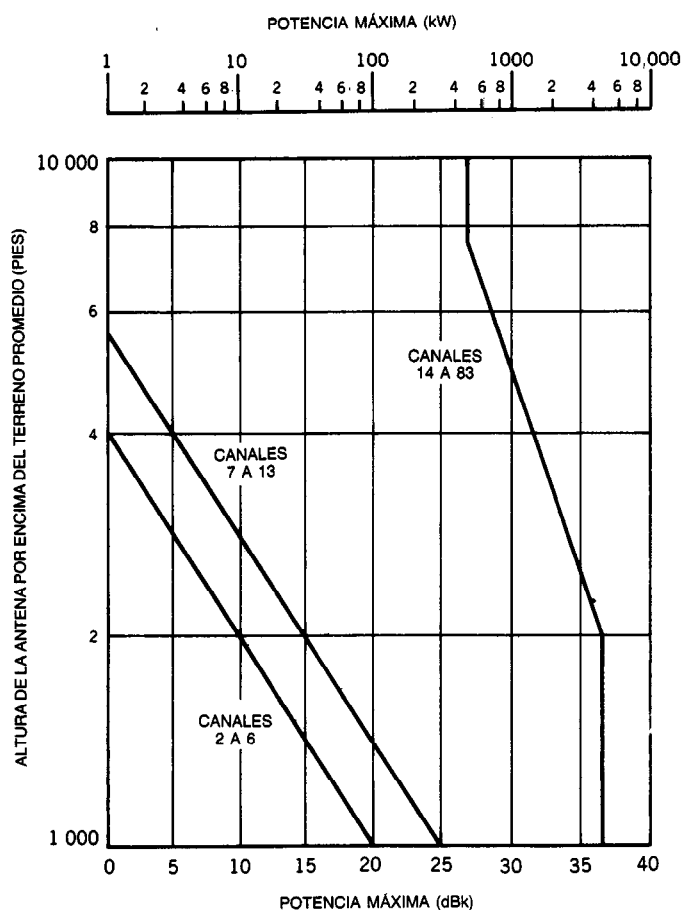


Fig. 49-28. Limitaciones de la altura de antena y potencia para la zona I.



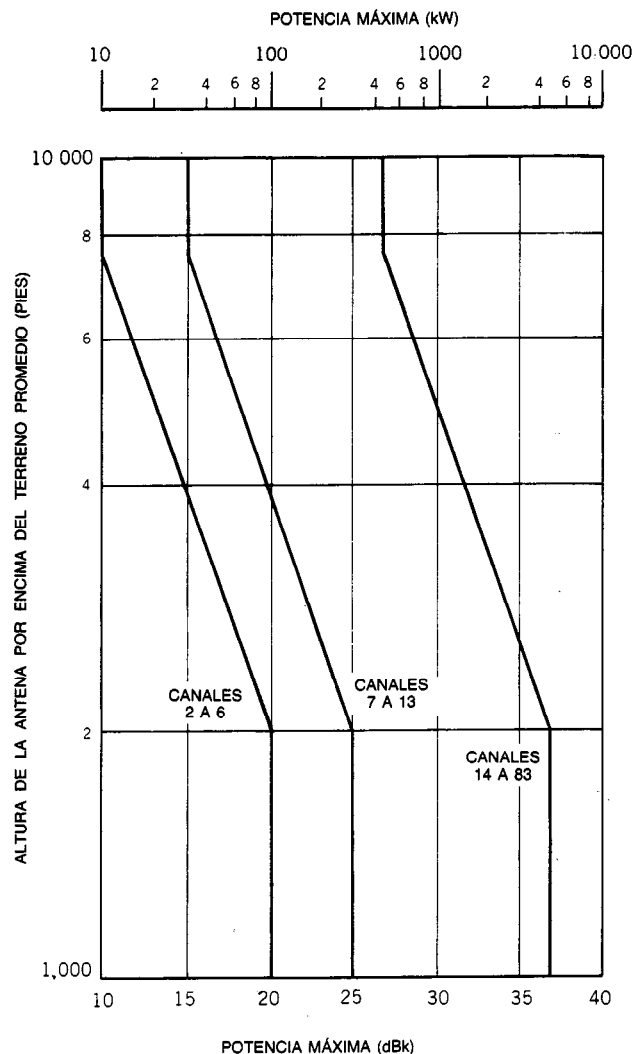


Fig. 49-29. Limitaciones de la altura de antena y potencia para las zonas II y III.

#### Sistemas de antena

En la generalidad de los casos, los sistemas de antena para televisión se diseñan a fin de proporcionar cobertura omnidireccional en el plano horizontal. Se requiere que la polarización tenga una componente horizontal; sin embargo, el empleo de la polarización circular se está haciendo más popular debido a que es posible obtener mayor cobertura, especialmente por la reducción de desvanecimiento de polarización y el aumento en potencia radiada efectiva. Las ganancias de potencia varían de 6 a 20 dB, dependiendo de la frecuencia y número de elementos. Los elementos pueden estar compuestos de dipolos, ranuras y hélices, así como de otros tipos menos conocidos. Dos factores de suma importancia en el diseño son la altura de la torre (de la antena) y la carga debida al viento. La FCC regula la potencia radiada efectiva para alturas de an-

tena por encima de 1 000 pies para estaciones VHF en la zona I y 2 000 pies para estaciones UHF de la zona I. En las figuras 49-28 y 49-29 se muestran estas limitaciones.

En la parte 73.609 del documento *Rules and Regulations* de la FCC se proporcionan descripciones detalladas de las zonas.

#### LIBROS Y REVISTAS

- A. F. Barghausen, "Medium Frequency Sky Wave Propagation in Middle and Low Latitudes", *IEEE Trans. Broad.* 12:1-14.
- G. W. Bartlett, ed., *National Association of Broadcasters Engineering Handbook*, 6.<sup>a</sup> ed., The National Association of Broadcasters, Washington DC.

- H. Bremmer, *Terrestrial Radio Waves: Theory of Propagation*, Elsevier, Amsterdam.
- Electronic Industries Association, Standard TR-101A, *Electrical Performance Standards for AM Broadcast Transmitters*.
- Federal Communications Commission, *Rules and Regulations*, Vol. III, Partes 73 y 74.
- "NTSC Signal Specifications" y "NTSC Color Standards", *Proc. IRE*.

### 49.3 REDES DE COMUNICACIONES

Surya V. Varanasi\*

#### 49.3.1 Introducción

Las comunicaciones de datos son la parte de las telecomunicaciones de desarrollo más rápido. Una razón primordial de ese rápido avance es el uso creciente del procesamiento electrónico de datos y de computadoras. Las redes de datos constan ya sea de una red de computadoras o de un conjunto de terminales conectadas a una o más computadoras, y son capaces de manipular información en muchas formas, incluyendo palabras escritas, símbolos, gráficas o simplemente sucesiones de bits. Las reservaciones aéreas y para ferrocarril, operaciones bancarias, control del tránsito aéreo y composición remota de textos para periódicos y revistas son solamente unos cuantos ejemplos de aplicaciones y empleo de sistemas de redes de comunicación de datos en la vida diaria actual.

#### 49.3.2 Consideraciones para el diseño de redes

La mayor parte de las redes de comunicación constan de nodos, anfitriones, terminales y enlaces de transmisión. Un nodo es una computadora cuya función principal es la conmutación de datos. Los anfitriones son computadoras que se utilizan esencialmente para funciones diferentes a la conmutación de datos. Los terminales son dispositivos de vinculación entre el usuario y la red, y los enlaces de transmisión reúnen este conjunto de elementos de subred a fin de formar una red de comunicación. Una red de datos consta de enlaces de transmisión, nodos y los programas de control esenciales.

Existen dos formas generales de organización de una red de datos: centralizada y distribuida. En la figura 49-30 se muestra una red básica de comunicación de datos de dos terminales. En las figuras 49-31 y 49-32 se presentan una red de datos centralizada y una distribuida, respectivamente. En una red de datos centralizada con una sola ubicación de procesamiento principal, todo el tráfico ocurre entre las terminales remotas y la única unidad de procesamiento central (CPU, de *central processing unit*). Por el contrario, en una red de datos distribuida, las capacidades más im-

portantes para el procesamiento de los datos se encuentran localizadas en más de una ubicación de procesamiento central.

En el diseño de una red de datos, a fin de satisfacer la demanda de la red, se establece una delicada solución entre las capacidades y el costo. Una red diseñada adecuadamente debe proporcionar comunicación confiable, libre de errores, en un tiempo razonable. Algunas consideraciones para el diseño de una red de datos son:

1. Organización de la red.
2. Tarifas y estructuras de tarifas.
3. Confiabilidad.
4. Tipo de servicios de comunicación, es decir, líneas conmutadas, arrendadas o privadas, o bien una combinación de éstas.
5. Trayectorias de las líneas.
6. Tipos de equipo terminal utilizado en sitios remotos.
7. Protocolos, es decir, localización y tipos de procedimientos para el control de la comunicación.
8. Procedimientos para el control de errores.
9. Economía.
10. Acceso sencillo para el usuario.
11. Seguridad, según sea requerida por el usuario.

En el diseño también es necesario tener en cuenta los requisitos para la aplicación a las necesidades del usuario, además de especificaciones técnicas como:

1. Número y localización de los sitios de procesamiento.
2. Número y localizaciones de las terminales alejadas.
3. Tipos de operaciones por procesar.
4. Intensidades de tráfico para cada tipo de operación y cada tipo de terminal.
5. Urgencia de transmisión de la información (oportunidad).
6. Patrones de flujo de tráfico.
7. Proporciones aceptables de error.
8. Disponibilidad requerida del sistema.

No siempre es necesario que en un diseño rentable de una red de datos se incorporen todas las especificaciones técnicas que se acaban de enumerar. En vez de ello, podría contarse solamente con las esenciales para una aplicación específica. En la tabla 49-9 se presenta el costo relativo de transmitir un megabit de información gráfica en Estados Unidos, suponiendo equipo arrendado y utilización del servicio durante ocho horas por día laborable. El factor de distancia se toma como de 1 400 millas terrestres (2 200 km), y se establece un índice de costo de 100 para servicio de télex normal de 50 baud.

#### 49.3.3 Terminales de datos

Los siete tipos de terminales que se utilizan en la actualidad son:

\*El autor desea expresar su agradecimiento al Dr. Roger Freeman y al Dr. Udo Pooch por permitirle utilizar materiales de sus libros de texto.

Tabla 49-9. Costo de transmitir 1 MB de información gráfica a 1 400 millas

Medio <sup>a</sup>	Costo respecto al índice	Observaciones <sup>b</sup>
Telegrama	1 650	Tarifa diurna 30 bits/palabra, mensajes de 100 palabras
Carta nocturna	280	Entrega nocturna/30 bits palabra
Télex	100	50 b/s
Marcado directo a distancia (103 A)	11	Marcado directo a distancia, 300 b/s, diurna (automática)
Autodin	4	De uso completo durante horas hábiles (red militar de Estados Unidos)
Marcado directo a distancia (202)	1.75	Conjuntos de datos (automática), 2 000 b/s
Carta (correo aéreo)	1.15	30 bits/palabra, 250 palabras/página
Banda ancha de la Western Union	1.00	Dúplex, 2 400 b/s
WATS (servicio telefónico de zona amplia)	0.75	Servicio telefónico especial de la ATT, 200 b/s por día hábil
Línea arrendada (201)	0.28	Dúplex, 2 000 b/s, comercial
Línea arrendada (303)	0.11	Dúplex, 50 kb/s, comercial
Cinta DEC por correo	0.10	Correo aéreo, cinta de 2.5 Mb
Cinta IBM por correo	0.017	Correo aéreo, cinta de 100 Mb

<sup>a</sup> Los números entre paréntesis identifican modems del Bell System utilizados en el servicio indicado.

<sup>b</sup> Mb, megabits.

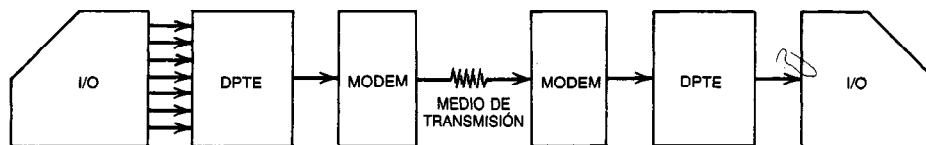


Fig. 49-30. Sistema básico de comunicaciones de datos de dos terminales; I/O = dispositivo de entrada-salida; DPTE = equipo de la terminal para el procesamiento de datos.

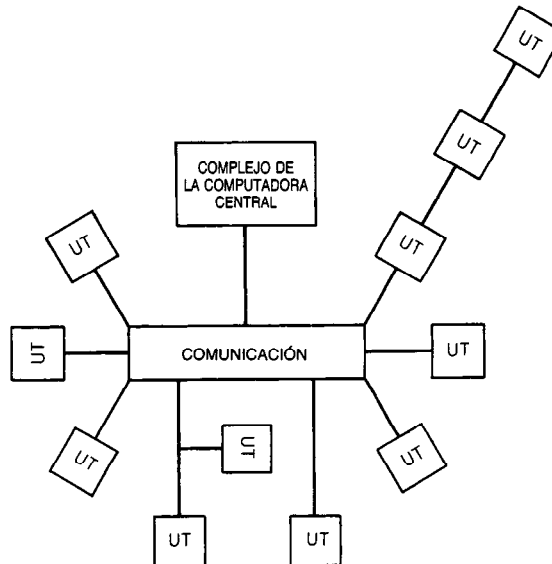


Fig. 49-31. Red de datos centralizada. UT = terminal del usuario.

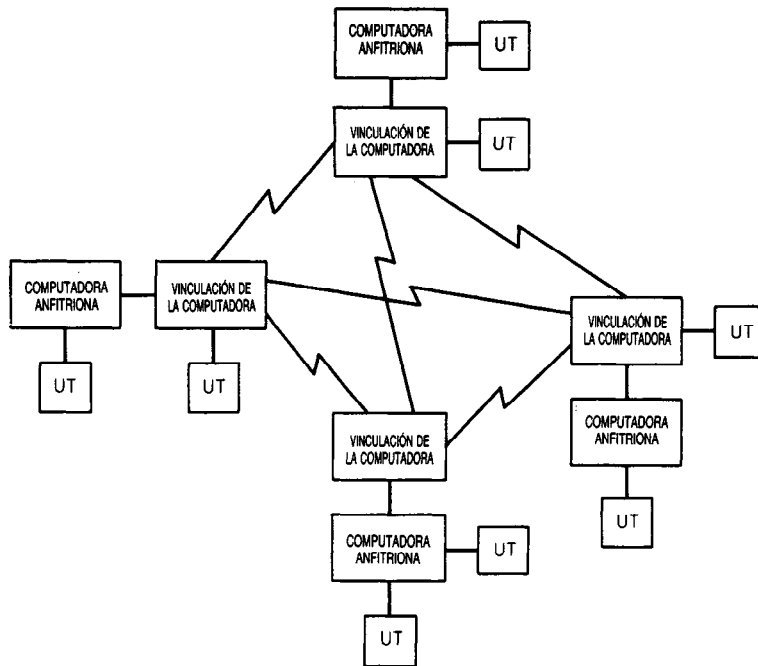


Fig. 49-32. Red de datos distribuida. UT = terminal del usuario.

1. Teleimpresora.
2. Tubo de rayos catódicos (CRT, de *cathode ray tube*) alfanumérico.
3. CRT gráfico.
4. Terminal para lotes remotos.
5. Terminal para preparación de datos.
6. Terminal para el sitio de ventas (POS, de *point of sale*).
7. Terminal para colección de datos de tipo industrial.

Una teleimpresora puede servir para entrada, salida, o ambas cosas. Las teleimpresoras pueden tener las siguientes configuraciones de fábrica:

1. Sólo de recepción (RO, de *receiving only*), que generan copias impresas.
2. Transmisora-receptora con teclado (KSR, de *keyboard send-recv*).
3. Transmisora-receptora automática (ASR, de *automatic send-recv*), con teclado, lectora de cinta de papel, perforadora de cinta de papel y una impresora de copias en papel.

Las teleimpresoras suelen ser asincrónicas (inicio-detención), y operar a una velocidad de 100 palabras por minuto. Los CRT alfanuméricos funcionan de forma parecida a las teleimpresoras, excepto que la impresora para copias en papel se sustituye por un CRT con las mismas capacidades. El CRT gráfico proporciona una visualización y entrada directa de representaciones gráficas de datos. Las terminales para manejo de lotes de datos remotos tienen medios de entrada y

salida (I/O), como lectora (y perforadora) de tarjetas, lectora de cinta de papel perforada, y unidades de cinta magnética a fin de tener acceso desde una terminal remota para procesamiento por lotes en la computadora. La línea de operación es una corriente de bits en serie a una velocidad de 1 200 a 9 600 b/s (bits por segundo). Este tipo de terminal suele denominarse "para entrada de tareas remotas" (RJE, de *remote job entry*).

La preparación de datos suele implicar la tradicional perforación de tarjetas, es decir, medios para la preparación de tarjetas de IBM, pero también es posible ampliarla desde el tablero a los sistemas de cinta o discos en los que el operador trabaja a partir de documentos fuente.

Un **sistema de sitio de ventas (POS)** es un sistema de terminal diseñado para su aplicación en establecimientos de servicio al público. Cada transacción se introduce directamente en una computadora o bien se coloca en almacenamiento fuera de línea para su introducción posterior. Una terminal POS parece una caja registradora mejorada. Otras entradas son lectoras de gafetes, de tarjetas de crédito y de código de barras. Muchas terminales POS cuentan con medios de procesamiento internos para calcular impuestos, efectuar operaciones aritméticas, comprobar dígitos de verificación, etc., y estas funciones pueden realizarse en línea o fuera de línea.

Las terminales para colección de datos industriales se encuentran esencialmente en áreas industriales, como almacenes, departamentos de recepción, estaciones (cuartos) de herramientas, líneas de producción y áreas de control de calidad (QC, de *quality control*).

La entrada de los datos se obtiene de forma óptica o magnética, o por medio de un teclado y tablero de lectura de tarjetas.

Las terminales de datos programables realizan funciones de vinculación, y están descargando a la CPU de ciertas funciones de preprocesamiento. La capacidad de programación se obtiene de un microprocesador integrado, que ayuda a que el usuario adquiera, edite, clasifique, actualice, archive, calcule y manipule datos fuente fuera de línea. La capacidad de memoria de una terminal es de 2 a 8 kilobytes.

Las terminales agrupadas son comunes; en esta modalidad todo un grupo de terminales (fig. 49-33) tiene acceso simultáneo al enlace de comunicación por medio de un controlador común del enlace de datos.

#### 49.3.4 Configuraciones de redes y conmutación de datos

En esta sección se analizan las redes de datos centralizadas y distribuidas desde el punto de vista del procesamiento de los datos en las redes, y se examinan las configuraciones de algunas redes de datos.

##### Configuraciones de redes

Una red de datos centralizada tiene estructura de "red en estrella", con una computadora central y las siguientes características:

1. Sus instalaciones de cálculo y conmutación se encuentran en un solo lugar.
2. Tiene el aspecto de árbol, aunque en algunos casos puede parecer un anillo o ciclo.

3. Existe una única trayectoria de comunicación entre la terminal y la CPU.
4. Es un sistema orientado a las terminales.
5. El tráfico fluye entre las terminales y la CPU.

En la figura 49-31 se muestra la configuración de una red de datos centralizada. En la red con configuración de anillo que se muestra en la figura 49-34, los puntos de entrada y salida de los datos forman una estructura anular. En una red de anillo, el tránsito puede ser unidireccional o bidireccional, en cuyo caso aumenta considerablemente la confiabilidad de la red. Los procesadores pueden ser sencillos o complicados. En la configuración sencilla, sólo un par de estaciones puede comunicarse al mismo tiempo, mientras que en las configuraciones complicadas, mediante el empleo de la multicanalización por división de tiempo (TDM, de *time division multiplexing*), los bits fluyen en ambas direcciones y todas las estaciones pueden intercomunicarse casi de forma simultánea.

Un **sistema multipunto** es aquel en el que dos o más terminales comparten una línea arrendada o privada. Una **conexión de punto a punto** es cualquier conexión entre un par fuente-destino. En las figuras 49-35 y 49-36 se ilustran estas configuraciones.

##### Conmutación de datos

Los dos métodos para la conmutación de datos son:

1. Conmutación de tráfico o circuital, también denominada conmutación de almacenamiento y retransmisión.
2. Conmutación por paquetes.

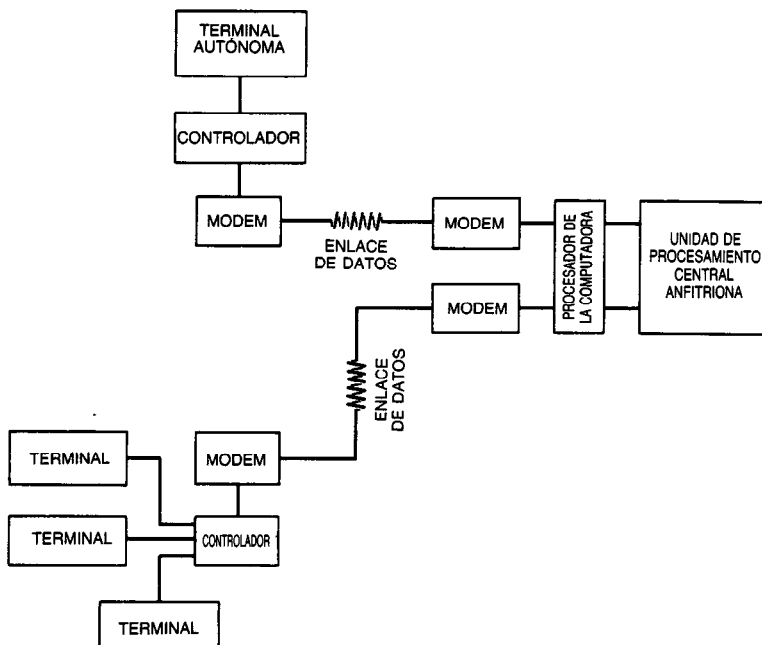


Fig. 49-33. Terminal unitaria o autónoma y cúmulo de terminales con acceso a una computadora anfitriona.

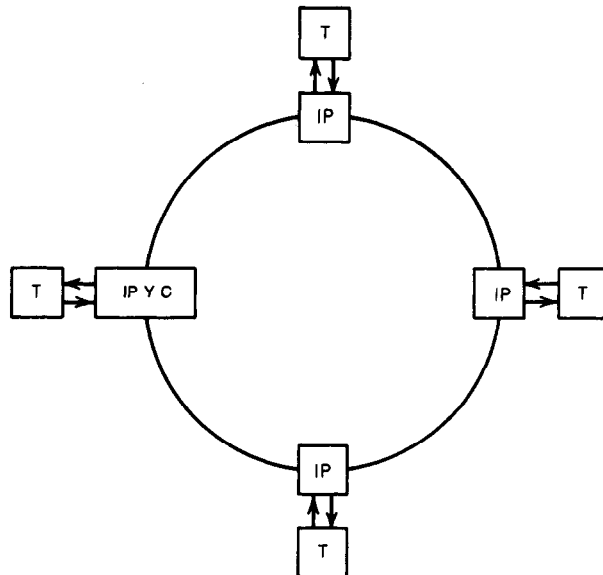


Fig. 49-34. Configuración de una red anular. IP = procesador de intercomunicación; T = terminal de datos o computadora; C = controlador de anillo.

**Conmutación de tráfico.** Para fines de comunicación, un **mensaje** puede definirse como una unidad lógica de información. Un mensaje está compuesto por:

1. Un encabezado que contiene información idónea para las operaciones de control de la red.
2. El "cuerpo" o texto, que contiene la información que será transferida.
3. Una "etiqueta (identificación) final", que contiene campos que significan el final del mensaje.

Algunos ejemplos de mensajes son los telegramas, programas y archivos de datos. Básicamente, en la conmutación de almacenamiento y retransmisión los datos se almacenan en nodos de conmutación, y ese tráfico se expide al nodo o dirección (o direcciones) siguiente(s) cuando queda disponible un circuito. Dentro de una red de conmutación de tráfico los mensajes se transfieren entre conmutadores en una base de mensaje por mensaje. Esto significa que un mensaje, algunas veces descompuesto en bloques de datos, debe ser transmitido y recibido por completo o bien cancelado a través de un enlace antes de que pueda transmitirse el siguiente mensaje. Cada bloque de un mensaje debe transmitirse en su sucesión idónea, de forma que el conmutador receptor sea capaz de reconstruir el mensaje y comunicar al conmutador emisor que ya lo ha recibido. En la figura 49-37 se muestra un formato típico de una red de conmutación de tráfico. Obsérvese que, en la figura, sólo el bloque inicial tiene suficiente información de control para continuar la transferencia. Un ejemplo de red de conmutación de tráfico es "Cybernet", diseñada e implementada por la Control Data Corporation (CDC). Algunas de las desventajas de la conmutación de tráfico son las siguientes:

1. Costos de conmutación elevados.
2. Grandes demoras del mensaje.
3. Menor eficiencia en la utilización de los recursos de la red.
4. Menos flexibilidad para ajustarse a las condiciones de tráfico.

Sin embargo, para un usuario regular, la conmutación de tráfico puede ser menos costosa que la conmutación circuital, ya que los costos de los circuitos se dividen entre los usuarios que comparten el sistema.

**Conmutación por paquetes.** En este tipo de conmutación, un mensaje de datos se descompone en partes denominadas paquetes, que pueden denominarse pequeños mensajes cortos, cada uno con su propio en-

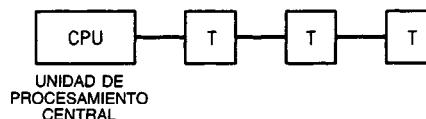


Fig. 49-35. Sistema de red de puntos múltiples. T = terminal.

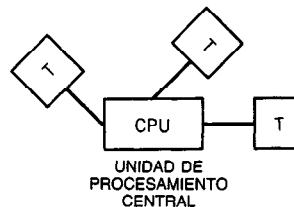


Fig. 49-36. Sistema de red de un solo punto. T = terminal.

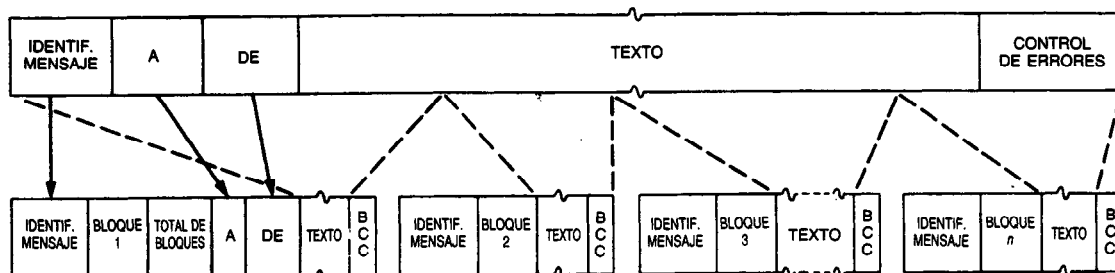


Fig. 49-37. Formato típico para una red de conmutación de mensajes. Tomado de Pooch y colaboradores, reimpresso con permiso. IDENTIF. = identificación. BCC = carácter de comprobación del mensaje.

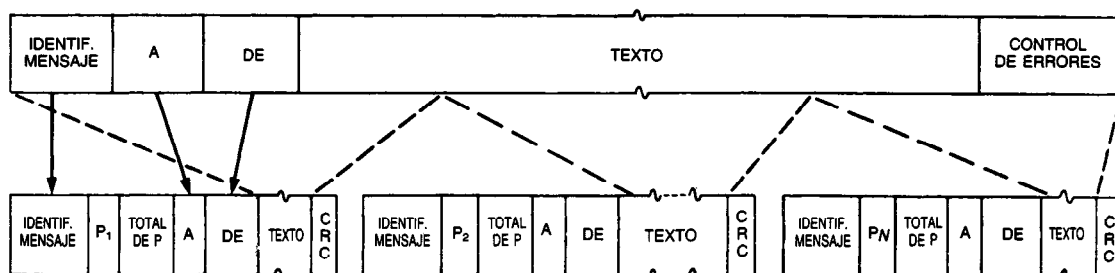


Fig. 49-38. Formato típico para una red de conmutación por paquetes. Tomado de Pooch y colaboradores, reimpresso con permiso. IDENTIF. = identificación. P = paquete; CRC = comprobación de redundancias cíclicas.

cabezado. A diferencia de lo que ocurre en la conmutación de tráfico, aquí cada paquete contiene suficiente información de control para su transmisión por una red, independiente de todos los demás paquetes que pertenecen al mismo mensaje. En la figura 49-38 se muestra el formato de la conmutación por paquetes. En la obra de Kleinrock se hace una descripción más detallada de una red de conmutación por paquetes. Una red de conmutación por paquetes bien conocida y popular es la red de la agencia de proyectos de investigación avanzada (ARPANET, de *advanced research projects agency network*).

#### 49.3.5 Optimización de circuitos y transferencia efectiva de datos

##### Optimización de circuitos

Las redes de datos, ya sean de conmutación por paquetes o de conmutación de tráfico, tienen pares fuente-destino cuyos requisitos varían con respecto a la "cantidad" de datos y su urgencia. No sería económico subutilizar líneas costosas que pueden aceptar 2 400, 4 800 o 9 600 b/s si la urgencia no fuese un factor crítico en una aplicación específica. De este modo, se requiere algún método con objeto de optimizar los enlaces para la comunicación de datos. Cuando las terminales están conectadas a un sistema de computación se utilizan dos técnicas básicas para la transmisión de datos: la multicanalización y la concentración.

**Multicanalización.** En este proceso, en un extremo del canal de comunicación de datos se combinan subcanales de velocidad inferior (es decir, corrientes de da-

tos de baja velocidad), y datos de velocidad superior se desmulticanalizan en los subcanales originales de velocidad inferior. Dos de las técnicas de multicanalización son la multicanalización por división de frecuencia (FDM, de *frequency division multiplexing*) y la multicanalización por división de tiempo (TDM, de *time division multiplexing*). En la FDM, el ancho de banda de las líneas de comunicación se subdivide en canales individuales más estrechos, evitando la intervenculación o interferencia entre canales. En la TDM se utiliza una señal de sincronización temporal a fin de asignar un segmento de tiempo específico a cada canal. La velocidad a la que los datos pueden entrar o salir de las terminales, cuando se multicanalizan, no puede exceder la velocidad de transferencia de datos del canal de comunicación.

**Concentración de datos.** Aunque es parecida a la multicanalización, conjunta datos de baja velocidad en caracteres y bloques con conversión de códigos, verificación de errores y compresión adecuadas. Más exactamente, los datos se combinan en una señal compuesta compleja, que contiene más información por unidad de tiempo. En el otro extremo de la línea de comunicación de alta velocidad, los datos se desconcentran en un procesador de entrada de forma parecida a como ocurre en el proceso de desmulticanalización. Gracias a la compresión de datos, se hacen más económicas las líneas de comunicación de alta velocidad.

##### Transferencia efectiva de datos

Supóngase que en una gran red de datos se estableciera un enlace de la fuente al destino, que funcionara a

2 400 b/s. ¿Cuán eficiente es el circuito? A esta velocidad, en una hora es posible entregar 8.64 Mb ( $2\,400 \times 60 \times 60$  bits). Pero la cuestión es cuántos bits en un periodo de tiempo son realmente útiles para la CPU en el destino. El American National Standards Institute recomienda el empleo del término "velocidad de transferencia de bits de información" (TRIB, de *transfer rate of information bits*) para designar la velocidad neta de transferencia de datos. La TRIB se define como el cociente del número de bits de información aceptados por el destino entre el tiempo total requerido para obtener los bits aceptados. La fórmula utilizada por Doll para hallar esta velocidad de transferencia de bits, suponiendo una transmisión por bloques, es:

$$\text{TRIB} = \frac{K_1(M - C)}{N_t(M/R) + \Delta T}$$

donde  $K_1$  = bits de información por carácter

$M$  = longitud del bloque del mensaje, caracteres

$R$  = velocidad de transmisión de la línea, caracteres/s

$C$  = número promedio de caracteres no de información por bloque

$N_t$  = número promedio de transmisiones requeridas para la aceptación del bloque en el destino

$\Delta T$  = tiempo entre bloques, segundos

Si  $P$  es la probabilidad de tener que retransmitir un bloque, entonces  $N_t$  puede expresarse como se indica a continuación:

$$N_t = \frac{1}{1 - P}$$

Así:

$$\text{TRIB} = \frac{K_1(M - C)(1 - P)}{(M/R) + \Delta T}$$

En esta fórmula no se hace referencia directa a la proporción de errores, ya que está implicada en el término  $(1 - P)$ .

Se concluye que existe una longitud de bloque óptima (dado el rendimiento del canal en términos de errores) que optimizará la transferencia de datos, al menos hasta donde concierne a  $M$ , con todos los demás términos constantes. También debe observarse que el término  $C$  puede reducir bastante la TRIB. Otros retardos pueden ser:

1. El tiempo dedicado a marcar el número.
2. El retardo de propagación del canal del satélite.
3. El retardo de sincronización del modem.
4. El tipo de solicitud de repetición automática (ARQ, de *automatic repeat request*), ya sea de detención y espera o de marcha continua en un dúplex.

Un enlace de datos puede estar tan sobrecargado de ineficiencias que, por ejemplo, en un enlace de datos de 2 400 b/s solamente pueden permitirse 100 b/s de TRIB de la fuente al destino.

#### 49.3.6 Consideraciones operacionales

La estructura de puntos múltiples fue una de las primeras técnicas establecidas para reducir el costo de la transmisión de datos a terminales remotas. En vez de conectar una sola terminal a cada línea remota se conectan varias, cada una con una o más direcciones únicas a esa terminal, como se muestra en la figura 49-39. Así, el tráfico recibido en una terminal, pero no direccionado a esa terminal, se ignora. Además, las configuraciones de puntos múltiples suelen tener la capacidad de "direccionamiento en grupo en todas direcciones", en el cual más de un terminal en una línea puede recibir el mismo mensaje sin que el emisor deba transmitirlo más de una vez. Una forma de estructura de puntos múltiples, que se utiliza para iteraciones de

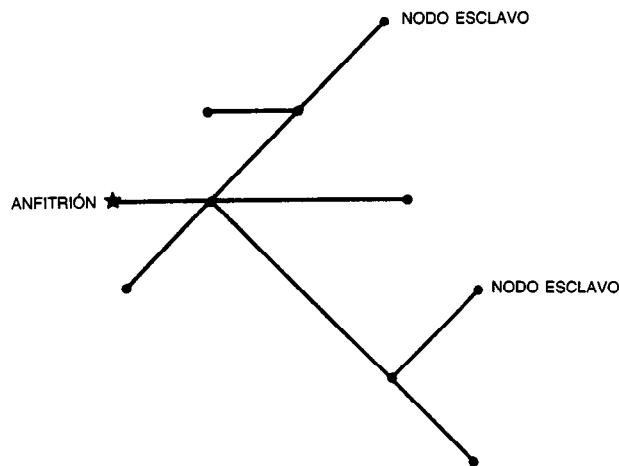


Fig. 49-39. Red de puntos múltiples. Tomado de Pooch y colaboradores, reimpresso con permiso.



corta distancia en las que no se requieren modems, implica el uso de una corriente de datos de transmisión sincrónica. Cada mensaje es precedido por la dirección de la terminal de destino; todas las demás terminales ignoran esta transmisión. Se transmiten pulsos de sincronización a fin de mantener sincronizadas las terminales. Las redes sincrónicas de puntos múltiples son muy populares cuando la detección de elementos omitidos en la red es esencial para su operación continua y eficiente. Otros procedimientos, a saber, la contención y la interrogación, se utilizan si más de una fuente de datos se encuentra conectada con la línea.

### Contención

Los procedimientos del protocolo de acceso con contención, que es popular debido a su sencillez, funcionan de forma adecuada si el usuario es bastante diestro en el manejo de las demandas hechas por la red. Si solamente hubiera dos o tres terminales, el funcionamiento podría darse sencillamente en una base de el primero en llegar, primero en despachar. Pero los esquemas más complicados requieren que la transmisión empiece solamente en instantes predeterminados, con lo que se reduce la probabilidad de interferencia.

El procedimiento de acceso con contención requiere que un mecanismo reconozca la recepción de una señal. Si no ocurre este reconocimiento, el emisor supone que los datos transmitidos fueron interferidos y retransmite hasta que ocurre tal reconocimiento. El acceso con contención es aplicable tanto a líneas terrestres como a medios de radio, pero es de más utilidad en radiocomunicaciones, en las que las restricciones de la asignación de frecuencias requieren que los usuarios compartan las frecuencias.

Un ejemplo de red bastante conocida por su empleo del procedimiento de acceso con contención es la red "Aloha", que conecta terminales en Hawái con varias localizaciones en el continente. En el sistema Aloha de radio por paquetes se utiliza un esquema de acceso con contención que permite a los usuarios la transmisión aleatoria de información en dos canales de 100 kHz asignados para ser utilizados por la red. Es posible lograr una utilización máxima del canal del 18.4% (Abramson), lo que representa una utilización deficiente de un recurso muy valioso: el espectro RF. El protocolo de acceso con contención Aloha ranurado constituye una mejora en cuanto a eficiencia respecto al procedimiento Aloha sencillo. En este esquema los paquetes pueden transmitirse sólo al inicio de un intervalo predefinido del reloj (Kleinrock y Lam). Cuando dos paquetes "tratan" de ocupar el canal al mismo tiempo, existe una colisión. La transmisión de paquetes solamente al inicio de un intervalo predefinido del reloj limita el tiempo en que pueden ocurrir las colisiones. Cuando ocurren, siempre lo hacen al inicio de un intervalo del reloj. Por consiguiente, las colisiones aleatorias no ocurren en una transmisión una vez que ésta ha empezado. El esquema de acceso con contención Aloha ranurado funciona de forma idónea si todas las terminales de la red están sincronizadas. Esto requiere un mecanismo de cronometraje muy preciso, así como un procedimiento para implementar la sincro-

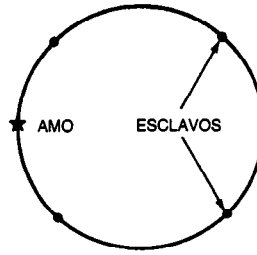


Fig. 49-40. Red de línea compartida por interrogación por turno de las estaciones. Tomado de Pooch y colaboradores, reimpreso con permiso.

nización. El máximo factor de utilización (usando el análisis de factor de carga) para un canal Aloha ranurado es del 36.8%.

### Interrogación

Este sistema implica el control central de todos los nodos de la red. En la red anular o de ciclo también puede utilizarse esta técnica, que básicamente es una operación de amo-esclavo(s) en la cual el amo pregunta a cada esclavo si tiene algo que decir. Si la respuesta es afirmativa, se da permiso al esclavo para transmitir o se le programa para que transmita más tarde. Actualmente se utilizan dos tipos de interrogación: por turno y por pase de lista.

**Interrogación por turno (o por ronda).** Con este tipo, una red de terminales sólo puede operar si la configuración de la red es idónea. Cada nodo esclavo se conecta en serie con otro nodo esclavo hasta que se ha completado una trayectoria de regreso al nodo amo, formando de esta manera una configuración cíclica ("ronda") como la que se muestra en la figura 49-40. El nodo amo inicia una interrogación al primer nodo esclavo del arreglo, que pasa el turno al siguiente nodo esclavo si no tiene nada que comunicar. Este proceso continúa hasta que ha tocado el turno a todos los nodos esclavos. El ciclo o ronda se rompe y debe volver a iniciarse cada vez que un nodo esclavo necesita comunicarse con el nodo amo. Siempre que falla un nodo esclavo del ciclo, éste se rompe y es necesario volver a configurarlo a fin de continuar la operación. La complejidad extra que se requiere para la recuperación de una de tales fallas debe ser compartida por cada nodo esclavo de la ronda, lo que incrementa el costo. Estas desventajas hacen que este procedimiento sea menos popular que otros esquemas, incluso siendo posible hacer ahorros en los circuitos para la comunicación.

**Interrogación "por pase de lista" (o simplemente "interrogación por lista").** Este concepto, a pesar de ser sencillo, agrega una dimensión de flexibilidad para satisfacer las demandas cambiantes más comunes de la mayor parte de los ambientes de procesamiento de datos. El nodo terminal solamente requiere "saber" que es un nodo esclavo de una red, y ser capaz de responder de forma adecuada cuando sea interrogado por el nodo amo. La terminal de pase de lista es considera-

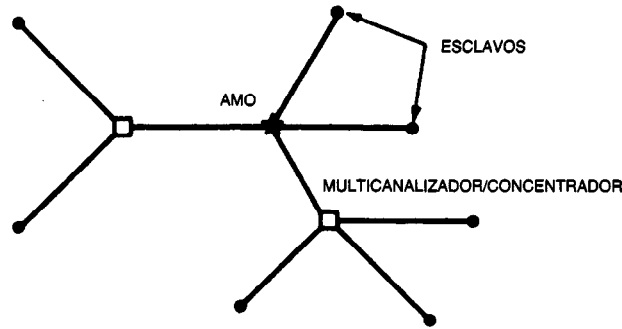


Fig. 49-41. Red de línea compartida por interrogación por pase de lista de las estaciones. Tomado de Pooch y colaboradores, reimpresso con permiso.

blemente menos complicada que la terminal de ronda, que también debe ser capaz de reconfigurarse a sí misma en caso de que un nodo esclavo adyacente falle. Sin embargo, la complejidad necesaria del amo puede incrementarse mucho, dependiendo del esquema de pase de lista específico que se utilice. En la figura 49-41 se muestra una red de pase de lista sencilla.

El procedimiento de pase de lista más sencillo es establecer una sucesión predefinida de nodos que aseguren que cada nodo esclavo tenga una oportunidad de comunicarse con el nodo amo, y sólo entonces proceder a otro ciclo. El intervalo de retardo entre transmisiones consecutivas es constante. Sin embargo, supóngase que algunos nodos requieren comunicaciones más frecuentemente que otros, o tal vez con menor frecuencia pero con prioridad más alta. Estas dos condiciones justifican una consideración especial mediante la asignación de más oportunidades para las comunicaciones que de otra forma serían comunes. Esto, por supuesto, reduce las oportunidades de comunicación de las terminales restantes, por lo que debe efectuarse un balance cuidadoso entre las necesidades de todos los usuarios y los recursos disponibles. Esta alternativa adicional permite a un nodo esclavo, por ejemplo el nodo A de la figura 49-42, pasar lista con mayor frecuencia que los demás nodos esclavos. Los cambios en el algoritmo para el pase de lista ocurren solamente en el nodo amo.

### 49.3.7 Protocolos

Los protocolos son herramientas comunes para controlar la transferencia de información entre sistemas de computación. En una red distribuida en la que participen tres o más computadoras, grandes o pequeñas, que se comuniquen a fin de transferir con exactitud un mensaje entre dos o más anfitriones, no solamente es necesario retener la integridad de la información de los anfitriones, sino que los mensajes deben ser correctamente formateados para su transmisión dentro de la red y después transferirse eficientemente a través de cada enlace a lo largo de la trayectoria elegida de la red. Cada uno de estos procesos requiere un procedimiento diferente y en la generalidad de los casos se

encuentra bajo el control de diversos elementos integrantes de la red. El control lógico de la asignación de recursos en una red complicada se logra por medio de los protocolos anfitrión a anfitrión, extremo a extremo y enlace a enlace, que se muestran en forma esquemática en la figura 49-43.

**Protocolos anfitrión a anfitrión.** También se denominan protocolos de transporte, debido a la función básica que realizan. Por ejemplo, considérese una base de datos que contiene elementos de información que deben ser transmitidos a una base de datos remota. El anfitrión de origen "sabe" la forma en que el destino debe recibir los datos, y configura de manera acorde los elementos de datos individuales. Sin embargo, el nodo de servicio del subsistema de comunicaciones podría confundirse si se le proporcionan los datos en esta forma básica. El anfitrión de origen debe reunir los datos, formatearlos en un mensaje y fijar la información adecuada del encabezado y la etiqueta final para que el nodo de servicio la interprete de manera correcta.

La información extra requerida para el encabezado y la etiqueta final varía de red a red. El encabezado consta esencialmente de aquellos campos que permitirán que el subsistema entregue el mensaje de forma precisa y a tiempo. Como mínimo, el anfitrión asigna campos de datos tales como los que se ilustran en la figura 49-44. La necesidad del campo de destino es evidente, pero el requisito acerca de la prioridad y clasificación de los campos variará dependiendo de las características de los datos transmitidos. El campo de precedencia permite que un usuario identifique la urgencia para entregar un mensaje, tal vez a expensas de

SUCESIÓN DEL PASE DE LISTA			
1	ABCDE	ABCDE	ABC ...
2	ABCDEABCDEA ...		
3	ABCDE ... BCDBEA ...		

Fig. 49-42. Algoritmo opcional de pase de lista. Tomado de Pooch y colaboradores, reimpresso con permiso.

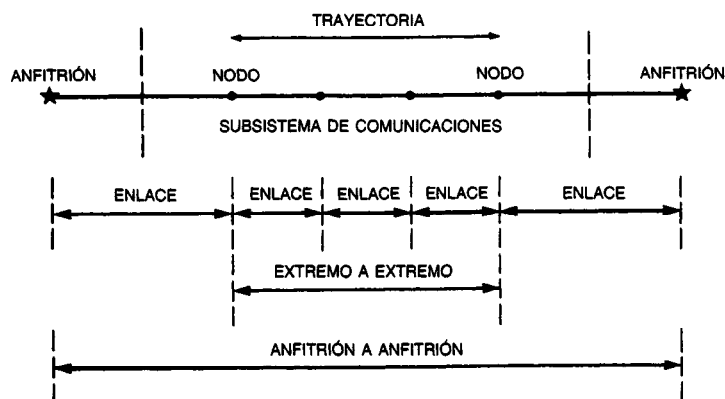


Fig. 49-43. Estructura jerárquica del protocolo. Tomado de Pooch y colaboradores, reimpresso con permiso.

retrasar otro tránsito de la red. El usuario puede querer que la entrega sea inhibida si inadvertidamente el anfitrión receptor no fue autorizado para recibir información patentada. Entonces, de alguna forma se notifica esta situación al emisor. El número del mensaje, aunque suficiente para algún intervalo de tiempo (por lo general, 24 horas), está limitado por el tamaño del campo. Sin embargo, si el mensaje está asociado con la fecha y hora de origen (DTG, de *date-time group*) y con el anfitrión de origen, entonces se encuentra iden-

tificado de forma única (y por siempre). El campo de origen se utiliza en el destino, de forma que pueda acusarse la recepción de datos y verificarse la totalidad de la entrada.

La información contenida en la etiqueta final suele ser básica. Consiste esencialmente en algún mecanismo para el control de errores y una indicación de que el texto ha terminado. Pueden incluirse o no campos extra, dependiendo de las necesidades específicas del ambiente.

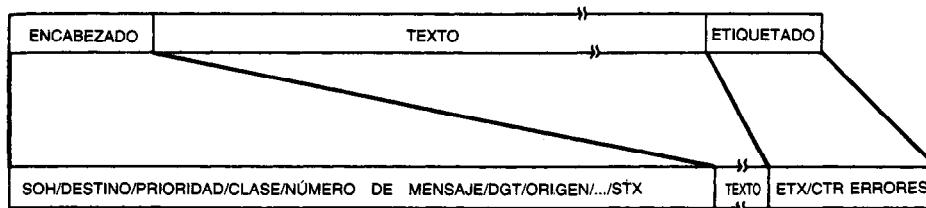


Fig. 49-44. Campo de datos proporcionado por un anfitrión. SOH = inicio del encabezado; DTG = fecha y hora de origen; STX = inicio del texto; ETX = fin del texto; CTR = control. Tomado de Pooch y colaboradores, reimpresso con permiso.

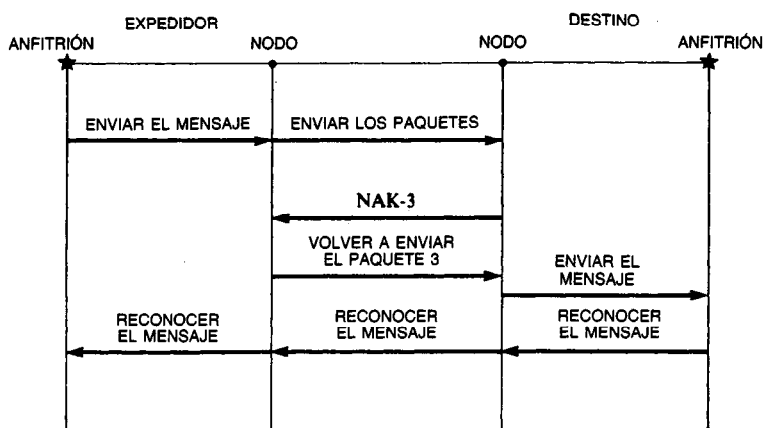


Fig. 49-45. Secuencia del protocolo de conmutación por paquetes de extremo a extremo. Tomado de Pooch y colaboradores, reimpresso con permiso.



**Protocolos enlace a enlace.** Permiten que los nodos adyacentes efectúen una comunicación organizada en circunstancias normales. Los protocolos de enlace, en el nivel inferior de la jerarquía de los protocolos, son los más complicados de todos los niveles. Los procedimientos de protocolo de enlace pueden dividirse en cuatro subprocedimientos: uno de las comunicaciones normales, otro para condiciones de falla de enlace "sin aviso", otro para la falla de enlace preadvertida, y uno más para el establecimiento del enlace. Cada uno de los procedimientos se subdivide aún más como se muestra en la figura 49-46.

El procedimiento para la falla de enlace preadvertida es sencillo. Los nodos vinculados a un enlace sólo deben advertir o reconocer que el enlace de interconexión se desactivará y, entonces, proceder a cerrarlo. Antes de la desactivación de un enlace, los nodos de conexión deben estar de acuerdo en el tiempo aproximado para intentar restablecerlo. A menudo, esta acción requiere la interacción del operador a fin de iniciar la sucesión adecuada de eventos en el instante correcto.

La sucesión para iniciar un enlace es un proceso fácil si en el primer intento se tiene éxito. En caso contrario, como a menudo sucede, el protocolo de enlace debe decidir entre volver a intentar el proceso ya iniciado o bien desactivar el enlace y notificar al operador. Esto suele estar basado en la cantidad de intentos fallidos realizados previamente. Luego de un número crítico predeterminado de intentos fallidos, no hay mucho por hacer de forma automática y debe entrar en acción el operador humano. Si se ha tenido éxito, el enlace continúa con el procedimiento normal del protocolo de comunicaciones.

Otro método para clasificar diferentes tipos de protocolos se basa en las técnicas utilizadas de encuadramiento de mensajes. Existen tipos de protocolos orientados a los caracteres, orientados a los bytes y orientados a los bits. Un protocolo típico orientado a los caracteres es el BISYNC o protocolo sincrónico binario de IBM. En el protocolo orientado a los caracteres se utilizan caracteres especiales para indicar eventos como "inicio del encabezado" (SOH, de *start of heading*), "inicio del texto" (STX, de *start of text*), y "final del texto" (EOT, de *end of text*). En el protocolo orientado a los bytes se utilizan delimitadores de la sucesión de caracteres en el encabezado, de forma parecida a como ocurre en el protocolo orientado a los caracteres, pero, además, se incluye un "contador" que tiene por objeto indicar el número de caracteres de datos en el mensaje.

Un tratamiento más detallado de los procedimientos y de los protocolos puede consultarse en la obra de Pouzin y Zimmermann.

#### LIBROS Y REVISTAS

- N. Abramson, "The Aloha System—Another Alternative for Computer Communications", *Proc. JICC*, pp. 281-285.
- D. R. Doll, *Data Communication Facilities, Networks and Systems Design*, Wiley, Nueva York.

- R. L. Freeman, *Telecommunications Systems Engineering, Analog and Digital Network Design*, Wiley, Nueva York.
- L. Kleinrock, "Analytical and Simulation Methods in Computer Network Design", *Proc. AFIPS JICC*, pp. 569-679.
- L. Kleinrock y S. Lam, "Packet Switching in a Slotted Satellite Channel", NCC, *AFIPS Conf. Proceed.* 42.
- U. W. Pooch, W. H. Green y G. G. Moss, *Telecommunications and Networking*, Little, Brown, Boston.
- L. Pouzin y H. Zimmermann, "A Tutorial on Protocols", *Proc. IEEE* 66(11):1346-1370.

## 49.4 SISTEMAS DE COMUNICACIONES POR SATÉLITE

*Daniel F. DiFonzo*

### 49.4.1 Introducción y repaso

Los satélites para comunicaciones, que actúan como repetidoras de radiofrecuencia (RF) en órbita, han revolucionado la industria de las telecomunicaciones desde su aparición en la década de 1960. La mayor parte de los satélites para comunicaciones se encuentran en órbitas geoestacionarias, que son trayectorias circulares en el plano ecuatorial a una altitud de 36 000 km. El periodo de la órbita es un día sideral, de modo que el satélite parece estacionario a un observador en la Tierra. Desde esta altitud es visible más de la tercera parte de la superficie terrestre. En la figura 49-47 se ilustra el empleo de un satélite para la retransmisión de señales a ubicaciones terrestres bastante separadas entre sí, A y B, por medio de enlaces directos de microondas en línea recta entre cada estación y el satélite. Un sistema de tres satélites geoestacionarios separados entre sí aproximadamente 120° en longitud proporcionaría una cobertura global e interconectividad total. Entre las características de los satélites para comunicaciones se incluyen un ancho de banda amplio con fines de elevada capacidad de canales para las comunicaciones del habla, así como para transmitir datos y TV digitales; acceso múltiple, de modo que un gran número de usuarios puedan comunicarse por medio del mismo satélite; tráfico con destinos múltiples y de punto a punto; transmisión directa a grandes zonas geográficas; comunicación con usuarios móviles, tales como barcos y aviones; enlaces entre satélites a fin de mejorar la interconectividad del sistema.

Un sistema de comunicaciones por satélite consta de una parte espacial y de una parte terrestre. La parte espacial incluye los satélites activos y de repuesto o reserva. La parte terrestre incluye las estaciones terrestres (que constituyen nodos de comunicaciones para el sistema) así como estaciones terrestres especiales para telemetría, seguimiento, comando y monitorización (TTC&M, de *telemetry, tracking, command, and monitoring*). Estas estaciones se utilizan para controlar la posición orbital del satélite, monitorizar su estado

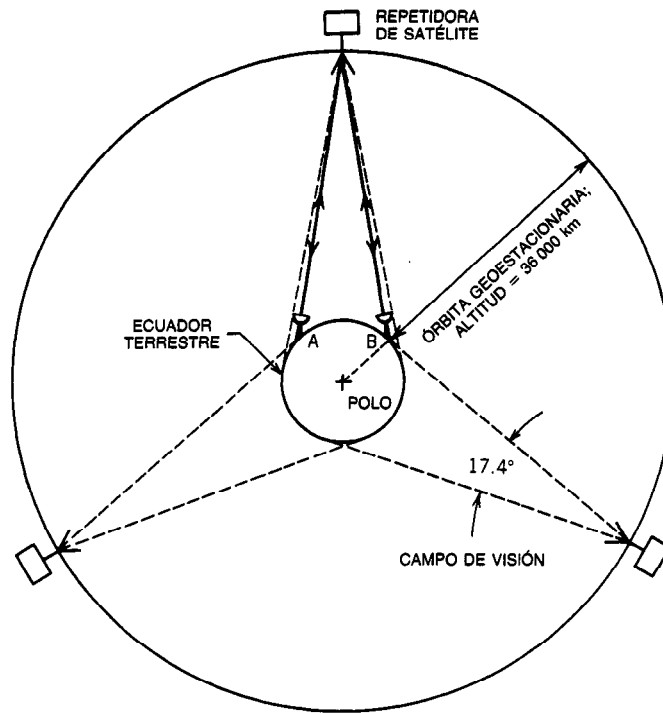


Fig. 49-47. Un satélite para comunicaciones es un repetidor de radiofrecuencias en órbita. Tres satélites geoestacionarios pueden cubrir la mayor parte de la superficie de la Tierra.

físico y volver a configurar sus circuitos en caso necesario. En algunos sistemas la parte terrestre también puede incluir redes de distribución terrestres en las interconexiones con otros sistemas de comunicación. En el sistema global INTELSAT, las instalaciones TTC&M se consideran integrantes de la parte espacial para efectos de asignación de costos.

En esta sección se describen la evolución y características de los sistemas de satélites para comunicaciones, así como la forma de la órbita y los vehículos de lanzamiento, los parámetros de los enlaces de comunicaciones, modulación y acceso, los subsistemas de las terminales de satélite y terrestres, sistemas de satélite representativos y los aspectos de coordinación.

#### Historia

En 1945, Arthur C. Clarke fue el primero en proponer satélites geoestacionarios para las comunicaciones por todo el mundo<sup>1</sup>. Describió las repetidoras "activas" o transpondedores, capaces de recibir señales de la Tierra y retransmitirlas. J. R. Pierce describió más tarde los conceptos de satélites pasivos y activos para la transmisión de señales<sup>2, 3</sup>. El lanzamiento del SPUTNIK por la Unión Soviética el 4 de octubre de 1957, y el del EXPLORER I por Estados Unidos el 1.º de enero de 1958, estimularon la actividad espacial. El 12 de agosto de 1960, el proyecto ECHO dio por resultado el lanzamiento de un globo metalizado hinchado en el espacio de 30 m de diámetro, para colocarlo en

una órbita terrestre con un apogeo (máxima altitud) de 1 688 km y un periodo orbital de aproximadamente dos horas. Este reflector pasivo podía retransmitir señales solamente entre antenas terrestres bastante separadas, mientras se encontraba en su campo de visión común. El 10 de julio de 1962 el TELSTAR, un satélite repetidor activo construido por los Bell Telephone Laboratories de la AT&T, fue lanzado por medio del vehículo Delta Thor de la NASA y puesto en órbita con un apogeo de 5 640 km, perigeo (mínima altitud) de 949 km, periodo de 158 minutos e inclinación de 44.8° con respecto al plano del ecuador. Pesaba 80 kg, su diámetro era menor de 1 m y podía recibir a 6 390 MHz y retransmitir a 4 170 MHz señales cuyo ancho de banda fuese de 50 MHz a un nivel de potencia del transmisor de 2 W. Se utilizó en la primera transmisión internacional en vivo (directo) por televisión<sup>3</sup>.

El periodo de una órbita geosincrónica (o sincrónica) es un múltiplo (por lo general igual a 1) del periodo de rotación de la Tierra. Si el plano de la órbita está inclinado con respecto al plano del ecuador, la trayectoria terrestre del satélite, o sea el punto subsatélite, delinea una figura en forma de ocho cuyo desplazamiento máximo en latitud norte-sur es igual a la inclinación orbital en grados. La órbita geoestacionaria ideal es un caso particular de la órbita geosincrónica con una inclinación de cero grados. Entonces, el punto subsatélite permanece estacionario sobre el ecuador. El SYNCOM I, construido por la Hughes Aircraft

Company para la NASA, fue lanzado el 14 de febrero de 1963, pero debido a la falla de un cohete no alcanzó la órbita que se pretendía. El SYNCOM II, lanzado el 26 de julio de 1963, logró una órbita sincrónica casi perfecta inclinada  $33^\circ$ , con apogeo y perigeo de 35 887 km (22 300 millas) y periodo de 23 horas 55.9 minutos. Recibió señales de enlace ascendente de la Tierra a 7 360 MHz y transmitió a 1 815 MHz. El SYNCOM III, lanzado el 19 de agosto de 1964 en una órbita geoestacionaria, transmitió con éxito la ceremonia de inauguración de los juegos de la XVIII Olimpiada desde Tokio hasta California<sup>4, 5</sup>.

La Communications Satellite Corporation (COMSAT) fue constituida el 1.º de febrero de 1963 para establecer comunicaciones sobre una base comercial. La COMSAT fue signataria de Estados Unidos y gerente principal de la International Telecommunications Satellite Organization (INTELSAT), creada el 20 de agosto de 1964. Originalmente, esta organización representaba 11 países que deseaban compartir la posesión del segmento espacial, que consistía en los satélites e instalaciones de la TTC&M, y que compartirían también la responsabilidad de los acuerdos sobre lanzamientos para un sistema mundial. Cada miembro de la INTELSAT cuenta con sus propias estaciones.

El satélite Early Bird (Pájaro Madrugador, denominado más tarde INTELSAT I), lanzado el 28 de junio de 1965, demostró la viabilidad comercial de un sistema de satélites geoestacionarios. Antes de esa fecha se había contemplado un sistema de satélites de mediana altitud (y por tanto no estacionarios), para el que se requería que las estaciones terrestres contaran con antenas capaces de seguir los satélites de horizonte a ho-

rizonte y "trasladarse" de un satélite a otro. Sin embargo, debido a que el sistema geoestacionario ha resultado mucho más rentable, se utilizan satélites de este tipo. El sistema INTELSAT ha crecido a tal punto que ya sostiene una proporción importante del tráfico internacional de telecomunicaciones. Actualmente la INTELSAT cuenta con 109 países miembros. Su segmento espacial consiste en 13 satélites en órbita geoestacionaria sobre los océanos Atlántico, Índico y Pacífico. El segmento terrestre incluye 834 antenas operacionales en 662 sitios, lo que representaba 172 usuarios. El sistema proporciona 1 318 rutas de comunicaciones entre estaciones terrestres portadoras de más de 35 000 circuitos de habla y datos, así como tráfico de TV.

INTELSAT no es la única entidad internacional de comunicaciones por satélite, ya que en 1971 se creó INTERSPUTNIK con fines de comunicaciones entre varios países del bloque del Este. La capacidad del segmento espacial se arrienda del satélite geoestacionario soviético GORIZONT<sup>6</sup>.

La International Maritime Satellite Organization (INMARSAT) se creó en 1979 para atender las comunicaciones móviles internacionales (p. ej., de barcos en el mar)<sup>7</sup>. También han aparecido muchos sistemas de satélites regionales y domésticos; hoy en día, existen más de 350 satélites activos, en reserva o planificados para la órbita geoestacionaria, y el mercado acumulativo para los satélites de comunicaciones es superior a 3 000 millones de dólares.

#### Repaso del sistema

En la figura 49-48 se muestran las características principales de un sistema de satélite repetidor de microon-

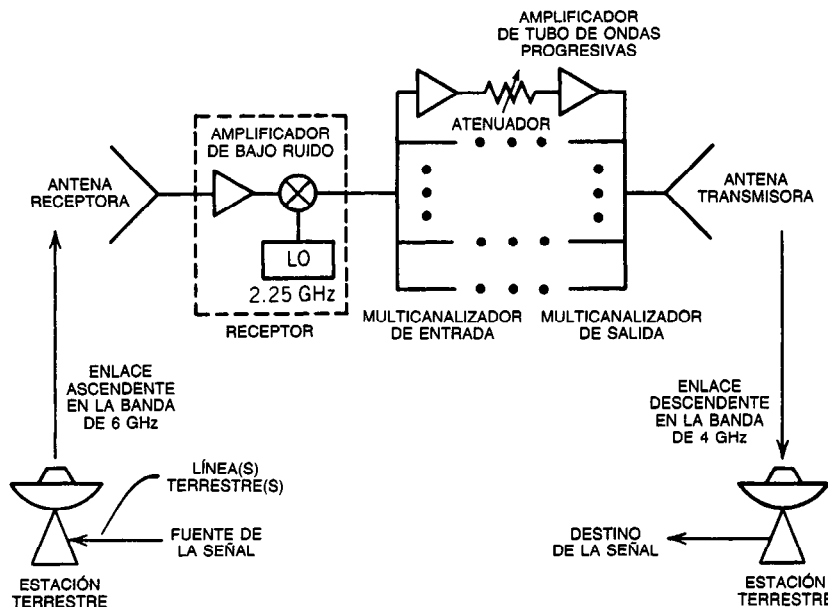


Fig. 49-48. Un satélite "de banda C" de acceso múltiple por división de frecuencia (FDMA), recibe señales de 6 GHz en el enlace ascendente y transmite señales de 4 GHz en el enlace descendente. LO = oscilador local.

Tabla 49-10. Lista parcial de las bandas de frecuencia para las comunicaciones por satélite

Frecuencia del enlace descendente (GHz)	Frecuencia del enlace ascendente (GHz)	Banda	Usos y sistemas comunes <sup>a</sup>
1.535-1.5425	1.635-1.645	L	Móvil marítimo, INMARSAT, MARISAT
2.5-2.655	2.655-2.69	S	FSS, INSAT móvil, ARABSAT (recepción de la radiodifusión para la comunidad)
3.4-4.8	5.85-7.075	C	FSS, INTELSAT, la generalidad de los satélites domésticos (pre-WARC 1979)
3.7-4.2	5.925-6.425		
4.5-4.8			
2.25-7.75	7.9-8.4	X	Móviles militares
10.7-12.7	12.75-13.25	K <sub>u</sub>	FSS, INTELSAT, SBS INTELSAT (pre-WARC 1979), SBS, ANIK
10.95-11.2			
11.45-11.7			
117-12.2	14.0-14.5	K <sub>a</sub>	Radiodifusión FSS, ACTS, INTELSAT, JAPAN experimental
12.2-12.7	17.3-17.7		
17.7-21.2	27.0-31.0		
40.0-41.0	50.0-51.0	V	FSS
41.0-43.0		Q	Radiodifusión
54.25-58.2			Enlace entre satélites
59.0-64.0			

<sup>a</sup> INMARSAT, International Maritime Satellite Organization; MARISAT, Maritime Satellite; FSS, Fixed Satellite Service; WARC, World Administrative Radio Conference; INTELSAT, International Telecommunications Satellite Organization; SBS, Satellite Business Systems; ACTS, Advanced Communications Technology Satellite.

das. Una señal de algún usuario terminal, tal como una llamada telefónica, ingresa en una estación terrestre por medio de las instalaciones terrestres. Ahí puede combinarse con otras señales, y estas señales modulan una portadora de radiofrecuencia de microondas de enlace ascendente, por ejemplo en 6 GHz. En la tabla 49-10 se presentan las bandas de frecuencia dominantes asignadas para las comunicaciones por satélite<sup>9</sup>. Los factores que afectan la elección de la radiofrecuencia serán analizados más tarde. Para el satélite de "banda C" que se ilustra en la figura 49-48, la señal del enlace ascendente en la banda de frecuencia de 5 925 a 6 425 GHz se recibe, amplifica y convierte (traslación lineal) a la banda del enlace descendente que va de 3.7 a 4.2 GHz. En la configuración de satélite de acceso múltiple por división de frecuencia (FDMA, de *frequency division multiple access*) que se muestra, el multicanalizador de entrada separa las señales por su frecuencia dentro de los canales individuales de transpondedor, donde se amplifican a un nivel adecuado para la transmisión en el enlace descendente. El multicanalizador de salida combina los canales individuales en una guía de ondas de salida común para la transmisión por medio de la antena del enlace descendente.

En la estación terrestre receptora, la señal se amplifica por medio de un receptor de bajo ruido, se convierte de manera descendente a una frecuencia intermedia, se desmulticanaliza y se envía al usuario terminal mediante instalaciones terrestres.

El sistema FDMA permite el acceso de muchos usuarios al mismo satélite sin interferencias mutuas por medio de la asignación de una única frecuencia a cada usuario. Los multicanalizadores del satélite permiten que cada transpondedor funcione independientemente de los otros. Otra técnica de acceso que cada vez está

adquiriendo más importancia es el acceso múltiple por división de tiempo (TDMA, de *time-division multiple access*), en el que los usuarios individuales comparten una asignación de frecuencia común, pero transmiten en pequeñas ráfagas conforme a sus segmentos de tiempo preasignados. Existe otra técnica de acceso, el acceso múltiple por división de códigos (CDMA, de *code-division multiple access*), en la que una única onda de forma se superpone a cada onda de forma del mensaje del usuario. En el CDMA, los usuarios comparten una banda común y los usuarios individuales se identifican mediante la descodificación de la onda de forma única. Ésta es una forma de comunicaciones de espectro extendido, y se aplica a sistemas militares y sistemas comerciales de baja potencia y baja velocidad de transferencia de datos.

En las comunicaciones por satélite se utilizan las modulaciones digital y analógica. La modulación analógica dominante es la modulación de frecuencia (FM), pero la tendencia es hacia las modulaciones digitales, incluyendo la manipulación por desfasamiento (PSK, de *phase shift keying*), la manipulación por desplazamiento de frecuencias (FSK, de *frequency shift keying*) y sus derivados. (Véase la sección 48.1.)

#### 49.4.2 Configuraciones geométricas de órbitas y lanzamientos

##### Órbitas

En las figuras 49-49 y 49-50 se ilustra una órbita elíptica general y un sistema de coordenadas geocéntrico. El centro de la Tierra se encuentra en uno de los focos de la elipse, *F*. Seis elementos describen el movimiento del satélite: excentricidad, *e*; longitud del semieje mayor, *a*; inclinación, *i*; ascensión recta del nodo ascen-



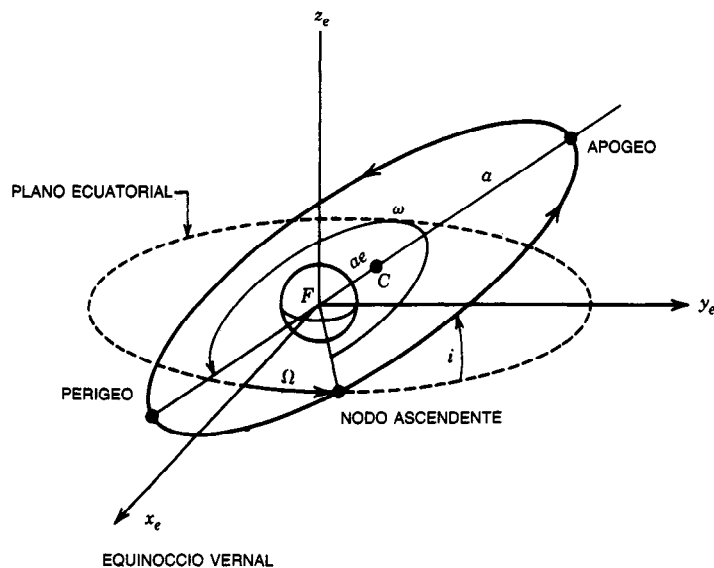


Fig. 49-49. Órbita elíptica en un sistema de coordenadas geocéntricas. El centro de la Tierra se encuentra en uno de los focos de la elipse,  $F$ . El centro de la elipse es  $C$ .

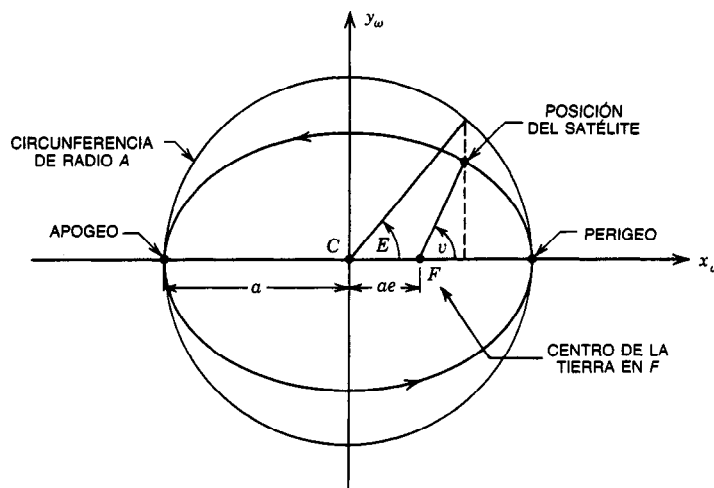


Fig. 49-50. Plano de órbitas ( $x_\omega, y_\omega$ ) en el que se muestran las anomalías verdadera  $v$  y excéntrica  $E$ .

dente,  $\Omega$ ; ángulo (en el plano de la órbita) del eje del perigeo,  $\omega$ ; y anomalía excéntrica (es decir, ángulo),  $E$ . La anomalía verdadera,  $v$ , es el ángulo real del eje de perigeo al satélite. La anomalía media,  $M = E - e \sin E$ , es el ángulo que se recorrería si el satélite se desplazara a su velocidad angular media. En la figura 49-50 se ilustra la anomalía verdadera  $v$  y la anomalía media  $E$ .

Miya<sup>10</sup>, Baker y Makenson<sup>11</sup> y Herrick<sup>12</sup> proporcionan fórmulas para calcular la posición del satélite a partir de los elementos orbitales. La ascensión recta y la declinación (ángulo norte con respecto al plano del

ecuador) del satélite pueden calcularse a partir de dichos elementos, y después relacionarse, por medio de fórmulas trigonométricas estándares, con la distancia y los ángulos de acimut y elevación para lugares específicos de la Tierra<sup>13,14</sup>. Sin embargo, debido a la naturaleza no ideal de las órbitas, el problema práctico de enfocar una estación terrena al satélite podría requerir fórmulas semiempíricas<sup>15</sup>.

La mayor parte de las cantidades de interés para las comunicaciones por satélite pueden obtenerse considerando el caso especial de órbitas circulares con inclinación cero. Entonces,  $e = 0$ ,  $I = 0$  y  $M = E = v$ .

Un satélite de masa  $m$  que se mueve a una velocidad  $v$  en una trayectoria circular con radio  $r$  (respecto al centro de la Tierra) experimenta una fuerza centrífuga  $mv^2/r$  igual y en dirección opuesta a la fuerza de atracción gravitacional  $GMm/r^2$ , donde  $GM = 398\,601.2 \text{ km}^3/\text{s}^2$  es el producto de la constante gravitacional  $G$  y la masa de la Tierra  $M$ , y se conoce con mayor exactitud que cualquiera de las cantidades por separado. Igualando las fuerzas, se obtiene:

$$v = (GM/r)^{1/2} = r(2\pi/T)$$

donde el periodo  $T$  de la órbita es el tiempo necesario para recorrer  $2\pi$  radianes. Para una órbita cuyo periodo sea un día sideral,  $T = 86\,164.091 \text{ s}$  y se encuentra que el radio de la órbita geosincrónica es  $42\,164.2 \text{ km}$ . El radio normalizado al de la Tierra ( $r_e = 6\,378.153 \text{ km}$  en el ecuador) es  $r_0 = r/r_e = 6.61072$ . En la tabla 49-11 se resumen las cantidades importantes para las órbitas circulares.

Para algunos cálculos de órbitas, por ejemplo para órbitas de transferencia, es útil conocer la velocidad y el periodo de una órbita elíptica. Dado un satélite a un radio  $r$  en órbita elíptica con semieje mayor  $a$ , la velocidad está dada por<sup>16</sup>:

$$v = \left[ 2 GM \left( \frac{1}{r} - \frac{1}{2a} \right) \right]^{1/2} \quad (49-1)$$

y el periodo es:

$$T = \frac{2\pi a^{3/2}}{\sqrt{GM}} \quad (49-2)$$

#### Relaciones geométricas y ángulos de dirección

En la figura 49-51 se presentan las relaciones geométricas de un satélite con respecto a una estación terrena en el punto  $p$  cuya latitud es  $\psi$  y su longitud este  $l_e$ . Para  $i = 0$ , la órbita normal es en la dirección  $+z_e$ . El satélite está en el eje  $x_e$  a una distancia normalizada  $r_0 = 6.6107$ . La cantidad  $\lambda$  es igual a  $l_e - l_s$ , donde  $l_s$  es la longitud este del punto subsatélite (directamente debajo del satélite). El arco  $\gamma$  es el ángulo esférico entre

el punto subsatélite y  $p$ . Las relaciones vectoriales entre las coordenadas centradas en la Tierra ( $x_e, y_e, z_e$ ) y las coordenadas centradas en la nave espacial ( $x_s, y_s, z_s$ ) producen la distancia normalizada  $S$ , y ángulos de dirección del eje del satélite a una estación terrestre:

$$S = |\mathbf{r}_s| = \sqrt{(1 + r_0^2 - 2r_0 \cos \psi \cos \lambda)} \quad (49-3)$$

$$\cos \theta_s = \frac{r_0 - \cos \psi \cos \lambda}{S} \quad (49-4)$$

$$\theta_s = \tan^{-1} \frac{\sqrt{(1 - \cos^2 \psi \cos^2 \lambda)}}{r_0 - \cos \psi \cos \lambda} \quad (49-5)$$

$$\phi_s = \tan^{-1} \frac{\sin \lambda}{\tan \psi} \quad (49-6)$$

$$\alpha_s = \tan^{-1} \frac{\cos \psi \sin \lambda}{(r_0 - \cos \psi \cos \lambda)} \quad (49-7a)$$

$$\alpha_s \approx \theta_s \sin \phi_s \quad (49-7b)$$

$$\beta_s = \sin^{-1} \frac{\sin \psi}{S} \quad (49-8a)$$

$$\beta_s \approx \theta_s \cos \phi_s \quad (49-8b)$$

donde  $(\theta_s, \phi_s)$  son los ángulos polares de las coordenadas del satélite y  $(\alpha_s, \beta_s)$  son los ángulos de dirección de acimut (este-oeste) y de elevación (norte-sur) del satélite a la estación terrestre.

En la figura 49-52 se ilustran los ángulos de acimut y de elevación de la antena de la estación terrestre con respecto al satélite,  $(\alpha_e, \beta_e)$ . El ángulo de acimut,  $\alpha_e$ , es el ángulo en sentido de giro a la derecha con respecto al norte que describe el gran arco de circunferencia  $\gamma$  con el meridiano local. Para determinar  $\alpha_e$ , primero debe calcularse la cantidad auxiliar  $A$ :

$$A = \tan^{-1} \frac{\tan \lambda}{\sin \psi}$$

Tabla 49-11. C cantidades para órbitas geoestacionarias circulares

Parámetro	Medida
Radio de la Tierra en el ecuador, $r_e$	$r_e = 6378.153 \text{ km}$
Masa de la Tierra, $M$	$M = 5.9734 \times 10^{24} \text{ kg}$
Constante gravitacional, $G$	$G = 6.673 \times 10^{-20} \text{ km}^3/\text{kg} \cdot \text{s}^2$
$G \times$ masa de la Tierra, $GM$	$GM = 3.986012 \times 10^{14} \text{ m}^3/\text{s}^2$
Radio de la órbita (plano del ecuador), $r$	$r = 42164.2 \text{ km}$
Radio de la órbita/radio de la Tierra, $r/r_e$	$r/r_e = r_0 = 6.61072$
Altura de la órbita en el ecuador, $h$	$h = 35786.04 \text{ km}$ $= 5.61072 \text{ radio de la Tierra}$
Velocidad, $v = \sqrt{GM/r}$	$v = 3.074662 \text{ km/s}$
Periodo, $T = 2\pi r^{3/2}/(GM)$	$T = 86164.091 \text{ s}$
Aceleración, $GM/r^2$	$GM/r^2 = 0.2242079 \times 10^{-3} \text{ km/s}^2$
Velocidad angular, $v/r$	$v/r = 72.92115 \times 10^{-6} \text{ rad/s}$



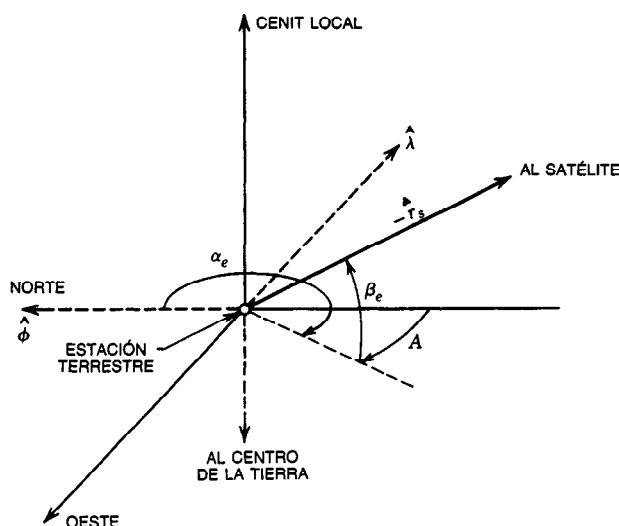


Fig. 49-52. Ángulos de acimut y elevación ( $\alpha_e$ ,  $\beta_e$ ) de una estación terrestre.

Tabla 49-12. Reglas para determinar el acimut

Latitud de la estación	Signo de $\lambda$ ( $\lambda = l_e - l_s$ )	Acimut ( $^\circ$ )
Norte	+	$180^\circ + A^a$
Norte	-	$180^\circ + A$
Sur	-	$A$
Sur	+	$360^\circ + A$

$$^a A = \tan^{-1} (\tan \lambda / \sin \psi).$$

atracción gravitacional entre el Sol y la Luna, y perturbaciones provocadas por maniobras de corrección. Una inclinación del plano de la órbita de  $i^\circ$  provoca que el punto subsatélite trace una figura en forma de ocho cada 24 horas, con declinación máxima (medida desde el centro de la Tierra) igual a la inclinación y el "ancho" angular del número 8 (es decir, la variación en la ascensión recta) de aproximadamente  $i^2/229^{\circ 13,14}$ . Desde una antena en una estación terrestre podrían verse variaciones semejantes, pero los ángulos exactos serían ligeramente diferentes de los de la ascensión recta y de la declinación, debido a una paralaje relativa al centro de la Tierra. Con base en la ecuación 49-2, podría deducirse que una reducción de 0.78 km en el radio de la órbita geoestacionaria con respecto a la ideal daría por resultado un desplazamiento hacia el este de  $0.01^\circ$  diario en longitud. Las fuerzas de atracción del Sol y la Luna afectan el radio, pero esencialmente perturban la inclinación del plano de la órbita. Estos efectos ocurren en un periodo de 18 años debido a la precesión lunar, además de una variación estacional cada 6 meses y un periodo de 14 días debido a que la órbita de la Luna no se encuentra en el plano del ecuador. De 1970 a 1988, la inclinación de la órbita varió entre  $0.73^\circ$  y  $0.93^\circ$  anuales, con una media de  $0.85^\circ$  por año<sup>19</sup> aproximadamente.

Las variaciones en longitud e inclinación debido a todas las causas requieren el encendido periódico de pequeños motores cohete de empuje a bordo del satélite a fin de mantener la longitud y orientación del plano de la órbita. La corrección de la longitud se denomina mantenimiento este-oeste de la estación, y la corrección de la inclinación, mantenimiento norte-sur de la estación. El encendido de los motores de empuje consume la mayor parte del combustible a bordo del satélite. Las correcciones de la órbita requieren cambios en el vector velocidad por medio de la adición vectorial de un incremento de velocidad,  $\Delta V$  ("delta  $v$ "). Para un satélite de masa  $m$ , incremento de velocidad  $dv$  y velocidad de escape relativa del combustible del motor de empuje  $v_e$  con dirección opuesta al vector velocidad del satélite, para la conservación de la cantidad de movimiento se requiere que  $mdv = -v_e dm$ . Para  $v_e$  constante, la masa relativa del combustible utilizado para obtener un incremento de velocidad  $\Delta V$  m/s es:

$$\frac{\Delta m}{m} = 1 - e^{(\Delta V/gI_{esp})} \quad (49-10)$$

donde  $g = 9.8 \text{ m/s}^2$  es la aceleración gravitacional al nivel del mar e  $I_{esp} = v_e/g$  es el impulso específico del motor de empuje, cuyas unidades son segundos.  $I_{esp}$  depende de los agentes químicos utilizados para el combustible; sus valores típicos varían de 220 a 300 s para la hidrazina y de 290 a 450 s para bipropulsores líquidos. Puede asumir valores muy grandes (p. ej., 2 500 s, aunque para un empuje bajo) para la propulsión iónica<sup>20</sup>. Son valores típicos:  $\Delta V \approx 54 \text{ m/s}$  por grado de cambio en la inclinación, y  $\Delta V \approx 2.8 \text{ m/s}$  por cambio de  $1^\circ$  diario en la velocidad de deriva en longitud. Los requisitos típicos para el mantenimiento este-oeste de la estación son de 3 a 6 m/s/año, y para el

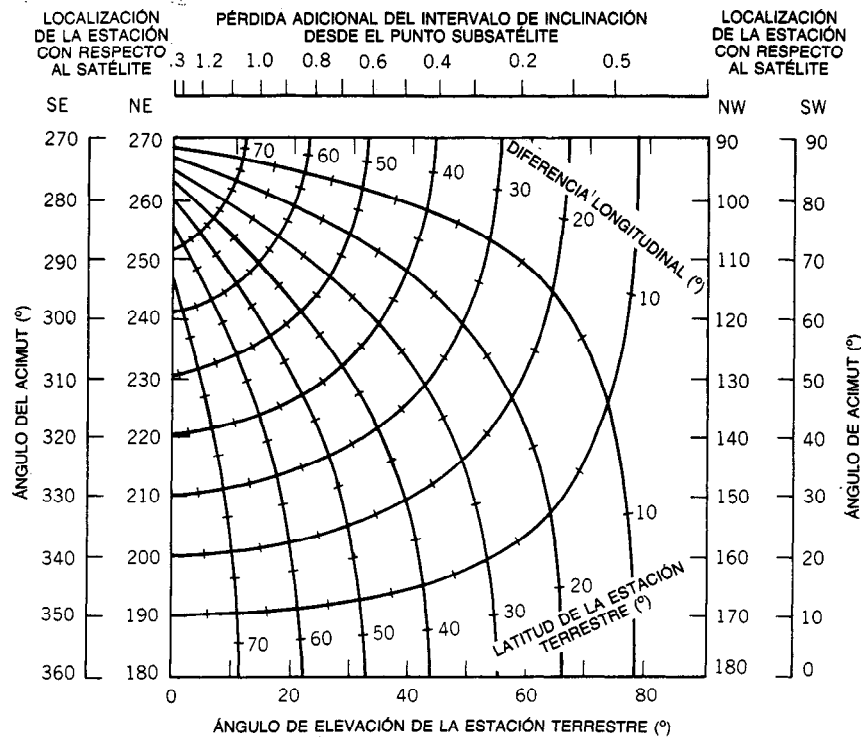


Fig. 49-53. Ángulos de acimut y elevación de una estación terrestre a un satélite geoestacionario. Cortesía de Communications Satellite Corp.

mantenimiento norte-sur, 50 m/s/año<sup>21</sup>. Por lo general, el mantenimiento este-oeste y norte-sur de satélites es exacto en  $\pm 0.1^\circ$ .

#### Orientación

En la figura 49-51 se esquematizan los ejes de balanceo ( $R$  de *roll*), cabeceo ( $P$  de *pitch*) y guiñada ( $Y$  de *yaw*) de un satélite. Los puntos en el eje de balanceo se "orientan" en la dirección de movimiento, los puntos en el eje de guiñada "caen" hacia el centro de la Tierra, y el eje de cabeceo está dirigido hacia el "sur" paralelo a la normal a la órbita. Estos ejes están rígidamente fijados al vehículo (nave) espacial. En el caso ideal están alineados con los ejes  $y_s$ ,  $-x_s$  y  $z_s$ , respectivamente, pero en la práctica se desvían de éstos por pequeñas cantidades. Si las direcciones del haz de la antena del satélite se mantienen fijas con respecto al cuerpo, entonces variará la dirección a la que señale el haz de la antena. En la figura 49-54 se muestra un cuadrante de la Tierra en unidades del ángulo de dirección del satélite, linealizado por medio de las aproximaciones de ángulo pequeño, ecuaciones 49-7b y 49-8b, que proporcionan los ángulos de dirección del satélite de acimut y elevación  $\alpha_s$  y  $\beta_s$ , respectivamente, para un punto  $p$ ; este punto parecerá en movimiento hacia el punto  $p_3$  debido a los errores en  $R$ ,  $P$  e  $Y$ . Un error de balanceo provoca que el haz se mueva en la dirección norte-sur, un error de cabeceo provoca un movimiento este-oeste, y un error de guiñada produce

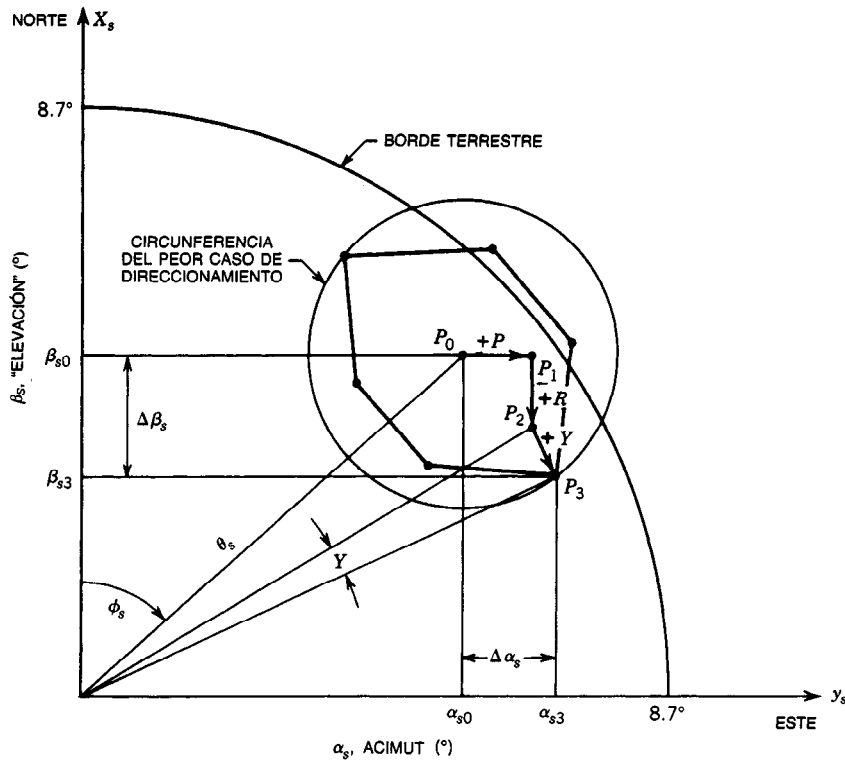
una rotación alrededor de los puntos subsatélite; por ejemplo, el campo eléctrico de una onda polarizada linealmente dirigida hacia el punto subsatélite giraría la cantidad del error de guiñada.

Para pequeños errores de  $R$ ,  $P$  e  $Y$ , expresados en grados, los cambios en los ángulos de dirección de acimut y elevación serían aproximados por:

$$|\Delta\alpha| \approx P + (Y/57.3)\beta_s \quad (49-11)$$

$$|\Delta\beta| \approx R + (Y/57.3)\alpha_s \quad (49-12)$$

La localización de  $p_3$  debe calcularse para todas las permutaciones de  $\pm R$ ,  $\pm P$ ,  $\pm Y$ , y el lugar geométrico de los puntos define el movimiento aparente de  $p$ . Cuando se requiere extrema precisión en el enfoque del haz, los desplazamientos exactos deben calcularse mediante el empleo de fórmulas estándares de trigonometría esférica. En la práctica, los errores no son independientes. Por ejemplo, los errores de guiñada y balanceo se transforman uno en otro cada seis horas (un cuarto del periodo orbital). Los errores tienen una parte determinista y una parte aleatoria, y los estadísticos exactos dependen del sistema específico para el control de la orientación o "actitud" utilizado por el satélite. Estos errores de orientación hacen necesario que las coberturas de la antena sean ensanchadas de forma adecuada, a fin de incluir todos los movimientos aparentes de las situaciones de cobertura terrestres.



**Fig. 49-54.** Mapa angular de la Tierra, vista desde un satélite. Los errores de orientación o "actitud" se han exagerado a fin de ilustrar el movimiento aparente de un punto  $p$  en la Tierra: desde  $p_0$  a  $p_3$  dados, (+) cabeceo seguido por (+) balanceo seguido por (+) guiñada. Los vértices del polígono corresponden a otras permutaciones.

### Control de la orientación

Los dos métodos dominantes para la estabilización son la estabilización de giro (*spin*) y la estabilización de tres ejes (fig. 49-55). En una nave (vehículo) con estabilización de giro se utiliza un tambor giratorio, en el cual los ejes de giro y cabeceo están alineados; la razón del momento de inercia alrededor de este eje entre el momento de inercia alrededor de cualquiera de los demás ejes suele ser de entre 1.1 y 1.3, y la velocidad de giro es de 100 r/min aproximadamente<sup>16</sup>. El "girador" tiene celdas solares (fotovoltaicas) alrededor de la superficie del tambor, y las antenas no deben girar de frente a la Tierra. En el tipo de giradores de "plataforma sin giro" de la Hughes, como el satélite INTEL-SAT VI, todo el paquete de RF está fijo (no gira) con respecto al tambor que contiene los demás subsistemas del satélite.

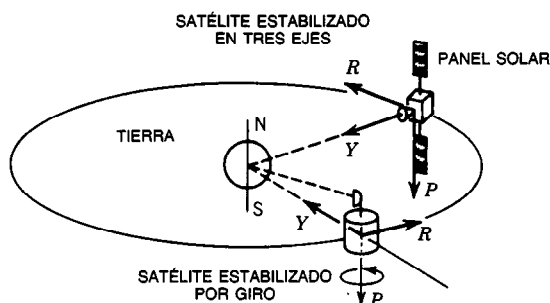
Los satélites de tres ejes mantienen la orientación por medio del intercambio de momentos torsionales entre varias ruedas de impulso ("volantes") que giran dentro del cuerpo del satélite. En la generalidad de los casos, las celdas solares tienen la forma de paneles que sobresalen de las caras norte y sur. Estos paneles deben girar a fin de estar de frente al Sol a medida que el satélite continúa desplazándose en su órbita. El satélite debe tener una rotación de cabeceo programada, a fin

de mantener las antenas dirigidas hacia la Tierra conforme el satélite se desplaza alrededor de ésta.

Para ambos tipos de satélites se utilizan detectores de infrarrojo para detectar el borde (limbo) de la Tierra contra el cielo frío. Además, se utilizan detectores de Sol y giroscopios integradores de viraje para afinar la determinación de la orientación o "actitud". El avance tecnológico mejora continuamente el control de orientación, y en la actualidad existe un debate relacionado con la estabilización de giro frente a la estabilización de tres ejes<sup>22</sup>. Los sistemas de control de la orientación actuales permiten que el balanceo y cabeceo se mantengan en  $\pm 0.1^\circ$ , y que la guiñada se mantenga a menos de  $0.25^\circ$  para ambos tipos de sistema. En un girador, los errores de cabeceo pueden corregirse por medio del intercambio de cantidad de movimiento sin consumo de combustible, pero las correcciones de balanceo y guiñada requieren el encendido de los motores de empuje. Para un satélite de tres ejes, el número y orientación de los volantes determinan si es necesario encender los motores cohete de empuje.

### Vehículos de lanzamiento y secuencia de lanzamiento

En la tabla 49-13 se resumen algunas de las capacidades de diversos vehículos de lanzamiento. Los vehículos "desechables" dominantes en la década de 1980



**Fig. 49-55.** Satélites estabilizados por giro y en tres ejes. Para el "girador", la antena de comunicaciones no debe girar con respecto a la superficie de la Tierra.

son los cohetes Delta<sup>23</sup> y Ariane<sup>24</sup>. Cada uno de ellos está disponible con una variedad de opciones. El Sistema de Transporte Espacial de la NASA (STS, de *Space Transportation System*) es un vehículo reutilizable que pone en órbita circular a una nave espacial, a una altitud de 296 km e inclinación de 28.5°<sup>25</sup>. En la figura 49-56 se ilustra la secuencia para el satélite INTELSAT VI. El STS se lanza a una órbita terrestre baja. El vehículo o nave es lanzado por el transbordador después de que se ha puesto a girar mecánicamente. Una vez que el vehículo se aleja a la deriva del STS, se orienta y se enciende su motor de perigeo a fin de colocarlo en la órbita de transferencia, que es una órbita elíptica con altitud de 296 × 35 786 km e inclinación de 27°. En el apogeo, se enciende el motor de apogeo (AKM, de *apogee kick motor*) para circularizar la órbita y aplicar suficiente  $\Delta V$  a fin de que la inclinación sea igual a 0°. A partir de la ecuación 49-1 y de la tabla 49-2, con  $r = r_e + 35\,786 = 42\,164$  km y  $a = (42\,164 + r_e + 296)/2 = 24\,419$  km, es fácil demostrar que la velocidad en el apogeo es 1 608 m/s. Dado que la velocidad geoestacionaria es 3 074 m/s (tabla 49-11), entonces se requiere un  $\Delta V = 1\,796$  m/s, que se obtiene por medio de la ley de los cosenos, para tomar en

cuenta la adición vectorial de velocidades y el cambio de 27° en la inclinación a fin de lograr la órbita estacionaria. Por lo general, el peso neto de la carga útil por colocar en una órbita geoestacionaria es aproximadamente el 58% del peso colocado en la órbita de transferencia. Los costos de los lanzamientos varían con el sistema de lanzamiento y dependen del año en que se realicen. El STS puede poner hasta 29 484 kg en su órbita estacionaria a 296 km; su bodega de carga útil mide 4.6 m de diámetro por 18.3 m de longitud. Los costos de lanzamiento del transbordador están en función de la fracción del peso y la longitud ocupada de la bodega de carga<sup>25</sup>. Típicamente, varios usuarios comparten la bodega de carga.

#### 49.4.3 Parámetros de enlace de los sistemas de comunicación por satélite

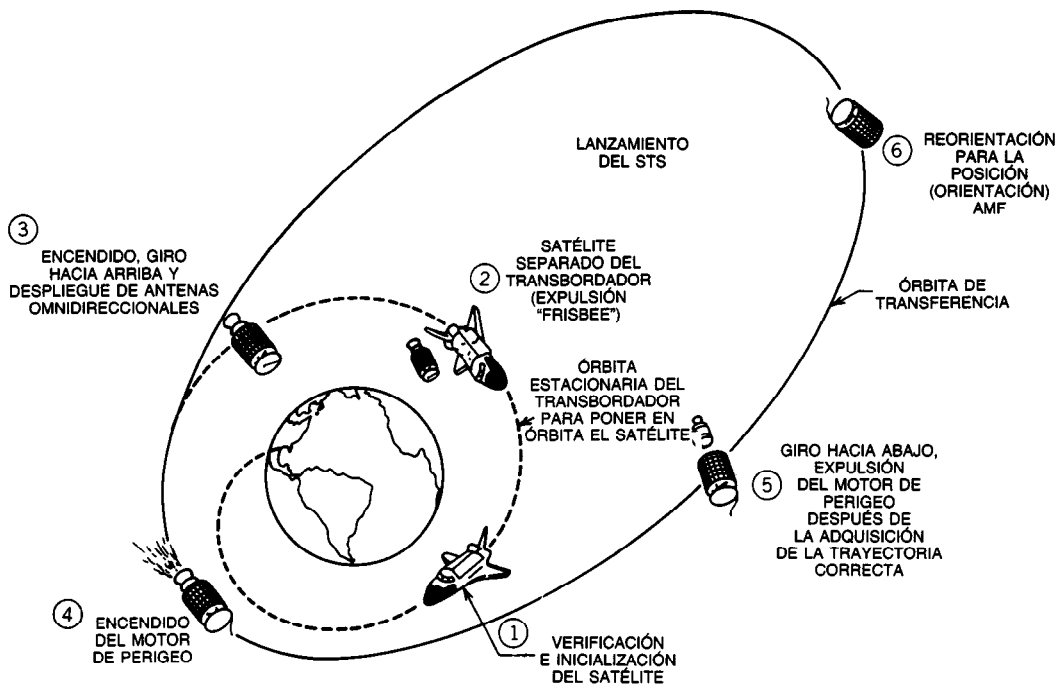
##### Ecuaciones de enlace

El punto de inicio para los cálculos de los enlaces de satélites es la relación de transferencia de potencia entre dos antenas separadas una distancia  $R$  (fig. 49-57). Las siguientes relaciones serán desarrolladas para un enlace en general, aunque es posible especializarlas

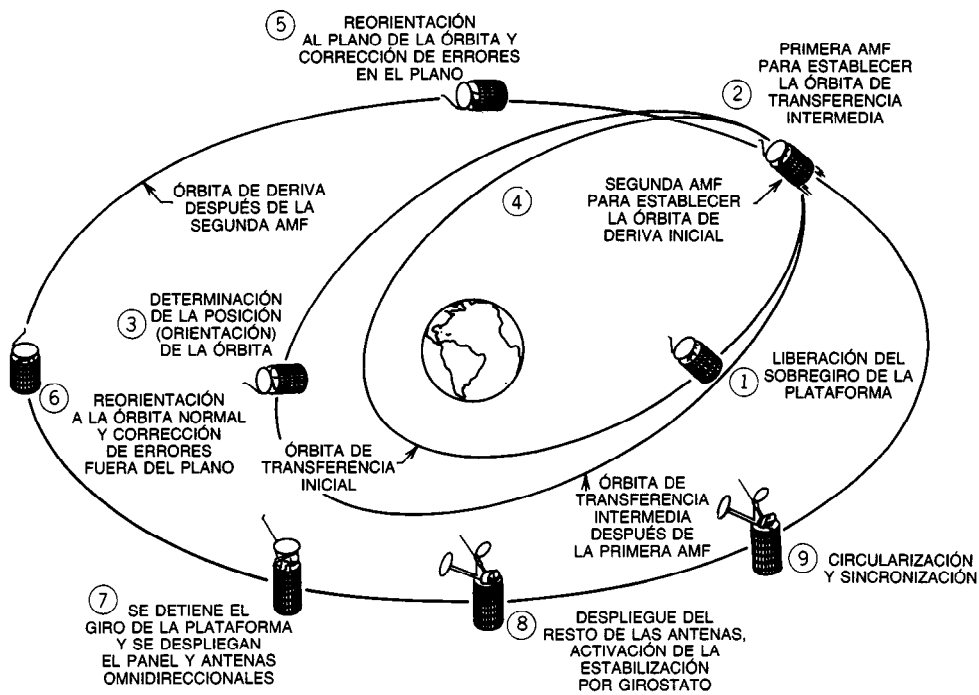
**Tabla 49-13. Vehículos lanzadores**

Vehículo <sup>a</sup>	Peso en la órbita de transferencia (kg)	Peso en la órbita geoestacionaria (kg)	Bodega contenedora de carga útil (mm)	
			Diámetro	Altura
Ariane 4 (ESA)	1852-4135	1080-2410	3650	8600
			3650	9600
Atlas-Centaur	2358	1093	2820	8540
Delta 3910/PAM	1111	600	2184	4368
Delta 3920/PAM	1247	680	2184	4368
N-II (Japón)		350	2200	
STS/PAM-D	1247	665	2184	2543
			2921	2543
STS/PAM-DII		983	2184	2543
			2921	2543

<sup>a</sup> ESA, European Space Agency; PAM, Payload Assist Module; STS, sistema de transporte espacial.



(a)



(b)

Fig. 49-56. Secuencia del lanzamiento del STS (sistema de transporte espacial) con el INTELSAT VI: a) lanzamiento y órbita de transferencia; b) órbita de transferencia a órbita geosincrónica. Cortesía de la Hughes Aircraft Co.



después para los casos del enlace ascendente y del descendente. Para una antena de transmisión con ganancia  $g_t$  (razón numérica) que radie  $p_t$  watts, la densidad de flujo de potencia a una distancia  $R$  es:

$$\psi = \frac{p_t g_t}{4\pi R^2} \cdot \frac{1}{l_i} \quad \text{W/m}^2 \quad (49-13)$$

donde  $l_i$  representa la atenuación atmosférica de la trayectoria debida a pérdida disipativa o a dispersión. Si una antena receptora de área efectiva  $A_r$  (no necesariamente igual a su área física) intercepta este flujo, la potencia recibida es:

$$p_r = \psi A_r \gamma \quad \text{W}$$

donde  $\gamma$  es el factor adimensional de acoplamiento de polarización ( $\gamma \leq 1$ ).

La potencia recibida no depende explícitamente de la frecuencia. Esta dependencia aparente se debe a la razón del área efectiva de una antena entre su ganancia, que es una constante universal para todos los tipos de antenas y está dada por:

$$\frac{A}{g} = \frac{\lambda^2}{4\pi} \quad (49-14)$$

donde  $\lambda$  es la longitud de onda. Para  $g = 1$ , esta razón es el área efectiva de una antena isotrópica,  $A_1$ , que disminuye con el cuadrado de la frecuencia, ya que  $\lambda = c/f$ , donde  $c = 2.998 \times 10^8$  m/s es la velocidad de la luz y  $f$  es la frecuencia en hertz. Su recíproco,  $4\pi/\lambda^2$ , es la ganancia de una antena con  $A = 1 \text{ m}^2$ . Para una antena recíproca pasiva, la ganancia, el área y el patrón de radiación no dependen del hecho de que la antena esté recibiendo o transmitiendo.

La calidad de un radioenlace depende de la razón de la potencia de la señal recibida,  $p_r$ , entre la potencia de ruido del sistema:

$$n = k T b_{\text{RF}} \quad \text{W}$$

donde  $k$  = constante de Boltzmann =  $1.38 \times 10^{-23}$  J/K

$b_{\text{RF}}$  = ancho de banda de radiofrecuencia, Hz

$T$  = temperatura de ruido del sistema, K

Entonces:

$$\frac{p_r}{n} = p_t g_t \cdot \frac{1}{4\pi R^2} \cdot \frac{\lambda^2}{4\pi} \cdot \frac{g_r}{T} \cdot \frac{1}{k} \cdot \frac{1}{b_{\text{RF}}} \cdot \gamma \cdot \frac{1}{l_i} \quad (49-15)$$

Se acostumbra expresar todas las cantidades en decibels. Entonces, la razón de portadora entre ruido (C/N) es  $C/N = 10 \log(p_r/n)$ :

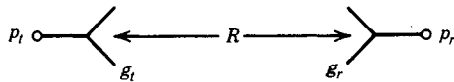


Fig. 49-57. Enlace RF entre dos antenas.

$$C/N = \text{EIRP} - L_s + A_1 + G_r/T + 228.6 - B + \Gamma - L_i \quad \text{dB} \quad (49-16)$$

donde cada término de la ecuación 49-16 es la expresión en decibels [es decir,  $10 \log(\quad)$ ] del término entre el paréntesis correspondiente. El término EIRP (dBW) se denomina potencia isotrópicamente radiada equivalente (de *equivalent isotropically radiated power*). Es la suma en decibels de la ganancia de la antena,  $10 \log(g_t)$  dBi (decibels con respecto a la isotropía), y la potencia de transmisión,  $10 \log(p_t)$  dBW.  $L_s$  (dB · m<sup>2</sup>) es el factor de dispersión en el espacio libre;  $A_1$  (dB · m<sup>2</sup>) es el área de una antena isotrópica;  $G_r/T$  (dB/K) es un factor de mérito para el sistema de recepción;  $-10 \log(k) = 228.6$  dBW/K · Hz;  $B$  (dBHz) es el ancho de banda de RF;  $\Gamma$  (dB) es el factor de polarización;  $L_i$  (dB) es la atenuación de la trayectoria.

Para una órbita geoestacionaria, la distancia de alcance,  $R = R_e$ , es entre 36 000 y 41 680 km, y el factor de dispersión correspondiente,  $L_s$ , varía entre 162.1 y 163.4 dB · m<sup>2</sup>, aunque por lo general se toma 163 dB · m<sup>2</sup> como típico para cálculos del enlace. Algunas veces se combinan  $L_s$  y  $A_1$  en lo que se denomina pérdida del espacio libre, que es un término dependiente de la frecuencia. Por ejemplo, a 4 GHz,  $\lambda = 0.075$  m, de forma que  $A_1 = -33.5$  dB · m<sup>2</sup> y la pérdida del espacio libre es  $L_s - A_1 = 196.5$  dB. El ancho de banda de RF  $B$  depende de los parámetros de modulación.

A fin de permitir la evaluación del enlace independientemente de este parámetro, se acostumbra normalizar la ecuación 49-16 al ancho de banda unitario. El resultado es la razón de portadora sobre densidad de ruido:

$$C/N_0 = (C/N) + 10 \log B \quad \text{dBHz}$$

Para la modulación digital, la cantidad de interés es la razón adimensional de energía por bit entre densidad de la potencia de ruido:

$$E_b/N_0 = (C/N_0) + 10 \log(t_b) \quad \text{dB} \quad (49-17)$$

donde  $t_b$  es la duración de 1 bit en segundos y  $1/t_b$  es la velocidad de transmisión de los bits en bits por segundo.

Dado que la generalidad de los parámetros de un enlace, como la frecuencia y el nivel de potencia, están determinados por restricciones del sistema, la relación  $(G/T)$ , que realmente se calcula como  $\log(g_r) - 10 \log(T)$ , es un factor de mérito que determina la calidad del enlace. Para una antena conectada con un receptor, la temperatura de ruido del sistema,  $T$ , es:

$$T = T_a + (l_g - 1)T_g + l_g T_r \quad \text{K} \quad (49-18)$$

donde  $T_a$  es la temperatura de ruido de la antena,  $l_g$  ( $\geq 1$ ) y  $T_g$  son el factor de pérdida y la temperatura física, respectivamente, de las partes no radiantes del sistema de alimentación y de la guía de ondas de conexión, y  $T$  es la temperatura de ruido equivalente del sistema receptor:

$$T_r = T_0 (f_r - 1) \quad (49-19)$$

Aquí  $T_0 = 290$  K y  $f_r$  es el factor de ruido expresado como una razón de potencias. La temperatura de la antena resulta de la potencia de ruido recibida de fuentes como la Tierra, el cielo, el Sol y la galaxia.

Lo anterior se aplica a cualquier enlace. Para una trayectoria de estación terrestre a estación terrestre vía satélite, las ecuaciones previas deben aplicarse por separado a los enlaces ascendente y descendente, y después se combinan los resultados.

La ganancia del transpondedor del satélite es un valor fijo diseñado con el objeto de activar el amplificador de salida a una potencia de salida de saturación específica para un valor particular de "densidad de flujo de saturación" del enlace ascendente. Algunos satélites tienen atenuadores controlables en la trayectoria de la señal, como se muestra en la figura 49-48, a fin de permitir la elección de varios valores posibles de densidad de flujo de saturación para suprimir el ruido del enlace ascendente. Los amplificadores de salida, ya sean tubos de ondas progresivas (TWT, de *traveling-wave tubes*) o amplificadores de potencia de estado sólido (SSPA, de *solid-state power amplifiers*), no presentan una característica de transferencia de entrada-salida lineal cerca de la saturación. En la figura 49-58 se muestra una curva de transferencia típica aplicable a muchos TWT, sin importar la potencia de salida absoluta. Una reducción en el nivel activador de entrada, denominada retroceso de entrada,  $BO_i$  (de *input back-off*), da por resultado un valor generalmente distinto de reducción en la potencia de salida, denominado retroceso de salida,  $BO_o$  (de *output back-off*). En la figura, un  $BO_i$  de 19 dB produce un  $BO_o$  de 12 dB. Las no linealidades generan productos de intermodulación cuando se encuentran presentes dos o más portadoras, y el elemento predominante es el producto de intermodulación de tercer orden. Esto incrementa el nivel de ruido y degrada la relación  $C/N_0$  global para el enlace. La definición de operación "lineal" es algo arbitraria, pero suele especificarse en términos de una razón dada de portadora entre potencia de ruido de intermodulación,  $(C/N)_{im}$ . Para un ancho de banda dada, un intervalo típico de valores aceptables de  $(C/N)_{im}$  es de 20 a 30 dB.

El nivel de activación afecta fuertemente la eficiencia de operación del transpondedor. En la figura 49-58 se implica que a fin de lograr  $(C/N)_{im} > 20$  dB para dos portadoras iguales,  $BO_i \approx 13$  dB y  $BO_o \approx 7.5$  dB. Por ejemplo, un TWT de 20 W solamente suministraría 3.5 W en total (7.5 dB por debajo de 20 W) o 1.75 W por portadora. Un tubo con eficiencia CD a RF del 35% en saturación funcionaría a una eficiencia sólo del 6%, ya que el consumo de potencia CD es fijo. Estas consideraciones alientan el empleo de esquemas de acceso de una sola portadora, como el TDMA, y el desarrollo de técnicas de linealización del amplificador. Debe observarse que por lo general el retroceso se aplica también a los transmisores de las estaciones terrestres. Si la estación debe transmitir muchas portadoras en todo el ancho de banda del satélite, el retroceso puede ser más sustancial que para el satélite.

La densidad de flujo del enlace ascendente, la ganancia de transferencia del transpondedor, la potencia de salida saturada del amplificador de salida, las pérdidas de salida, y la ganancia de antena del enlace descendente con el satélite determinan la EIRP del enlace descendente. La distancia de alcance del enlace descendente, la atenuación de la trayectoria y el factor de mérito ( $G_{re}/T_{re}$ ) de la estación terrestre se incorporan entonces a la ecuación 49-16 para calcular  $(C/N_0)_d$ , la razón de portadora sobre densidad de ruido del enlace descendente. La razón  $(C/N_0)_t$  del enlace global está relacionada con las contribuciones individuales por medio de:

$$\left(\frac{C}{N_0}\right)_t = \left[\left(\frac{C}{N_0}\right)_u^{-1} + \left(\frac{C}{N_0}\right)_{im}^{-1} + \left(\frac{C}{N_0}\right)_d^{-1}\right]^{-1} \quad (49-20)$$

Podría parecer que la  $C/N_0$  global puede incrementarse sencillamente maximizando la EIRP de la estación terrestre; sin embargo, la ecuación 49-20 indica que, para portadoras múltiples, existe un valor "óptimo" del retroceso del TWT del satélite que maximiza  $(C/N_0)_t$ . Esto limita el  $(C/N_0)_t$  máximo que es posible obtener y, con ello, la capacidad de comunicaciones del transpondedor<sup>26</sup>. La capacidad también depende de la modulación y del método de acceso. En las secciones 48.2 y 48.5, así como en las referencias 27 y 28, es posible encontrar ejemplos de los cálculos de enlaces.

#### Factores de rendimiento de los enlaces

**Frecuencia.** Para un enlace en el espacio libre, y con base en la ecuación 49-15, la pérdida de trayectoria total es:

$$\frac{P_r}{P_t} = k_1 \frac{g_r g_t}{f^2}$$

donde  $k_1$  incluye todos los demás parámetros. Si ambas antenas tienen fija la característica de ganancia frente a frecuencia, la pérdida aumenta con  $f^2$ , lo cual sugiere el empleo de la frecuencia práctica más baja. Si ambas tienen área de apertura fija, entonces el numerador es proporcional a  $f^4$  y la pérdida disminuye con  $f^2$ , lo que sugiere el empleo de la frecuencia práctica más elevada. Si una antena tiene ganancia constante y la otra tiene área efectiva constante, entonces la pérdida del enlace en el espacio libre es independiente de la frecuencia. Una antena con área física constante, tal como una de reflector o una red, no siempre tiene ganancia proporcional a  $f^2$ . En particular, las antenas de reflector con haz perfilado para satélite para la cobertura de una región fija de la Tierra tienden a presentar ganancia constante. Por supuesto, muchos otros factores afectan la elección de la banda de frecuencias para un enlace por satélite, incluyendo el ancho de banda, asignaciones de frecuencias en todo el orbe, interferencia, compartimiento de frecuencias con servicios terrestres, potencia disponible, avance tecnológico en los componentes de hardware, y factores de propagación.

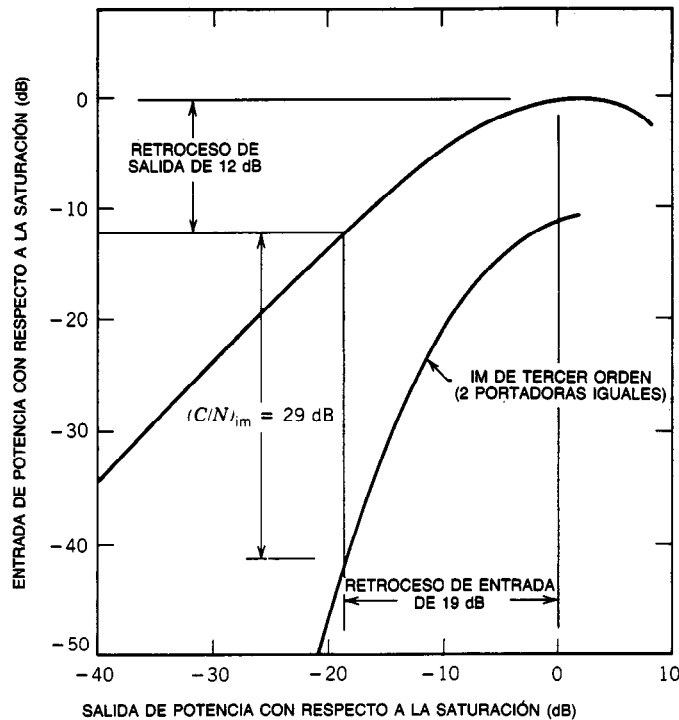


Fig. 49-58. La respuesta típica de transferencia de potencia de un tubo de ondas progresivas (TWT) no es lineal cerca de la saturación. IM = intermodulación.

**Propagación.** En la figura 49-59 se muestra la atenuación cenital unidireccional total a través de la atmósfera como función de la frecuencia (ref. 18, informe 205-4). Para otros ángulos de elevación,  $5^\circ \leq \beta_e \leq 90^\circ$ , es necesario multiplicar la ordenada por la cosec  $(\beta_e)$ , que es la razón aproximada de la longitud de la trayectoria en  $\beta_e$ , con respecto a la del cenit. En la figura 49-60 se observa la atenuación específica  $\gamma_R$  en decibels por kilómetro para algunas tasas de precipitación pluvial  $R$  como función de la frecuencia, con base en la relación exponencial:

$$\gamma_R = kR^\alpha \quad \text{dB/km}$$

donde las curvas son para valores de  $k$  y  $\alpha$  tanto para polarizaciones vertical como horizontal obtenidas del informe 721-1 del CCIR<sup>28</sup>, en el que se supone la distribución de Laws y Parsons<sup>29</sup> del tamaño de las gotas para gotas de lluvia esferoidales oblongas a  $20^\circ\text{C}$ . La atenuación real es  $A = \gamma_R L_s$ , donde  $L_s$  es la longitud de la trayectoria inclinada dada para  $\beta_e \geq 10^\circ$  como (ref. 28, informe 564-2):

$$L_s = \frac{5.1 - 2.15 \log[1 + 10^{(\psi-27)/5}]}{\sin \beta_e} \text{ km} \quad (49-21)$$

donde  $\psi$  es la latitud de la estación terrestre y  $h_0$  es su altitud (km). La atenuación excedida en un 0.01% respecto a la de un año promedio puede estimarse como:

$$A_{0.01} = \gamma_R L_s \left[ \frac{90}{90 + 4L_s \cos(\beta_e)} \right] \quad \text{dB} \quad (49-22)$$

Además de la atenuación, la atmósfera contribuye a la potencia del ruido recibido en una estación terrestre. Las contribuciones a este ruido incluyen fuentes galácticas (significativas por debajo de 1 GHz), y la absorción de señales por vapor de agua, oxígeno y los hidrómetros. En la figura 49-61 se muestra la temperatura de ruido atmosférico para humedad moderada ( $10 \text{ g/m}^3$  de vapor de agua), 1 atm de presión superficial, y temperatura superficial =  $20^\circ\text{C}$  (refs. 10; 28, informe 720-1). Por debajo de 1 GHz, el ruido galáctico aumenta con rapidez y domina el ruido atmosférico. Por consiguiente, existe un amplio mínimo en el intervalo de 1 a 10 GHz entre el ruido cósmico y el efecto de absorción/radiación del vapor de agua. Los componentes ortogonales de las ondas electromagnéticas que pasan a través de la lluvia experimentan una atenuación diferencial y defasamiento diferencial, como se muestra en las figuras 49-62 y 49-63<sup>30</sup>. La despolarización resultante de las ondas introduce componentes en polarización cruzada que pueden degradar el funcionamiento de los enlaces con reutilización de frecuencia mediante los cuales se intenta incrementar la capacidad a través del empleo simultáneo de polarizaciones ortogonales duales. El estadístico resultante de discriminación por polarización cruzada (XPD, de *cross-polarization discrimination*) depende de la longi-

tud efectiva de la trayectoria, de las distribuciones de tamaño de las gotas y del ángulo de inclinación de las gotas de lluvia con respecto a las polarizaciones incidentes<sup>31,32</sup>.

Otros efectos significativos de la propagación, descritos con mayor detalle en las referencias 10, 31 y 32, son: la refracción atmosférica, el centelleo (desvanecimiento rápido), la rotación de Faraday y el ruido debido a la lluvia.

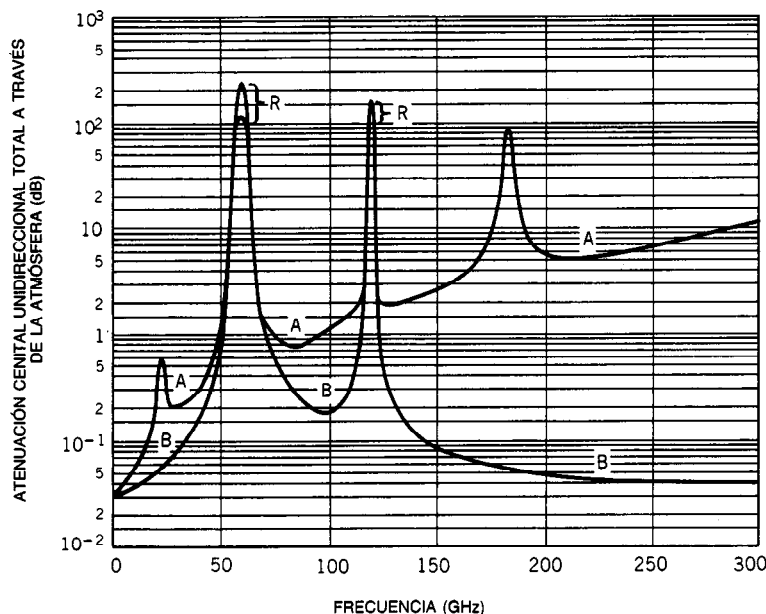
**Reutilización de frecuencia.** Los primeros satélites contaban con relativamente pocas placas (celdas) solares, y su capacidad de canales estaba limitada por la potencia de RF disponible. Satélites posteriores, empezando con el INTELSAT IV<sup>33</sup> a principios de la década de 1970, contaban con suficiente potencia y estaban "limitados en ancho de banda" en su capacidad de portar un mayor número de canales para el habla dentro del ancho de banda asignado de 500 MHz, la "banda C". A fin de lograr mayor capacidad del sistema es posible considerar varios métodos. Pueden buscarse nuevas asignaciones de frecuencias, pero esto no siempre es práctico, ya que las asignaciones dependen de acuerdos internacionales y rara vez son modificadas. Asimismo, la elevada inversión en las instalaciones existentes alienta el máximo aprovechamiento de las bandas de frecuencia existentes. Otro método es utilizar más de un satélite para un sistema dado, como se muestra en la figura 49-64. Cada una de las estaciones terrestres A y B tiene dos antenas operando en las mismas bandas de frecuencia, y las trayectorias son independientes mientras la potencia radiada por la an-

tena A en dirección del satélite B y viceversa sea aceptablemente baja (típicamente de 28 a 30 dB por debajo de la señal deseada). Los niveles de interferencia dependen de la separación del arco orbital  $\lambda$  entre los satélites, las envolventes de los lóbulos laterales de las antenas y las EIRP relativas de cada estación. Algunas veces este método se denomina reutilización de frecuencia en la Tierra. Para la generalidad de los sistemas, los lóbulos laterales de la antena de la estación terrestre deben mantener un nivel prescrito de máxima radiación a lo largo del arco geostacionario, definido por:

$$G(\theta) \leq \begin{cases} G_0 - 25 \log \theta & \text{dBi} \quad 1 \leq \theta \leq 48^\circ \\ -10 & \text{dBi} \quad \theta \geq 48^\circ \end{cases}$$

donde  $\theta$  es el ángulo fuera de eje en grados. El valor de  $G_0$  era 32 para los primeros sistemas, pero debido al poblamiento del arco geostacionario una regla reciente de la FCC (CC Docket 81-704) especificó  $G = 29$  a 6 GHz para todos los lóbulos laterales dentro de  $9^\circ$  del eje a lo largo del arco y  $G_0 = 19$  para la respuesta de polarización cruzada a 6 GHz. El objetivo de estas restricciones es permitir que alguna vez los satélites lleguen a estar tan cerca entre sí que solamente exista una separación de  $2^\circ$  sobre el arco geostacionario.

La capacidad del sistema de satélite también puede incrementarse mediante la reutilización de las bandas de frecuencia en el satélite. Esto puede lograrse por medio del empleo de polarizaciones ortogonales duales, así como de haces múltiples aislados espacialmente. En la figura 49-65 se muestran ambos conceptos,



**Fig. 49-59.** Atenuación cenital unidireccional total a través de la atmósfera. Curva A, humedad moderada (humedad de  $7.5 \text{ g/m}^3$  cerca del suelo). Curva B, atmósfera seca (humedad  $0 \text{ g/m}^3$ ).  $R$  = intervalo de los valores debido a la estructura fina. Tomado de Recommendations and Reports de la CCIR<sup>18</sup>, informe 208-4, con permiso.

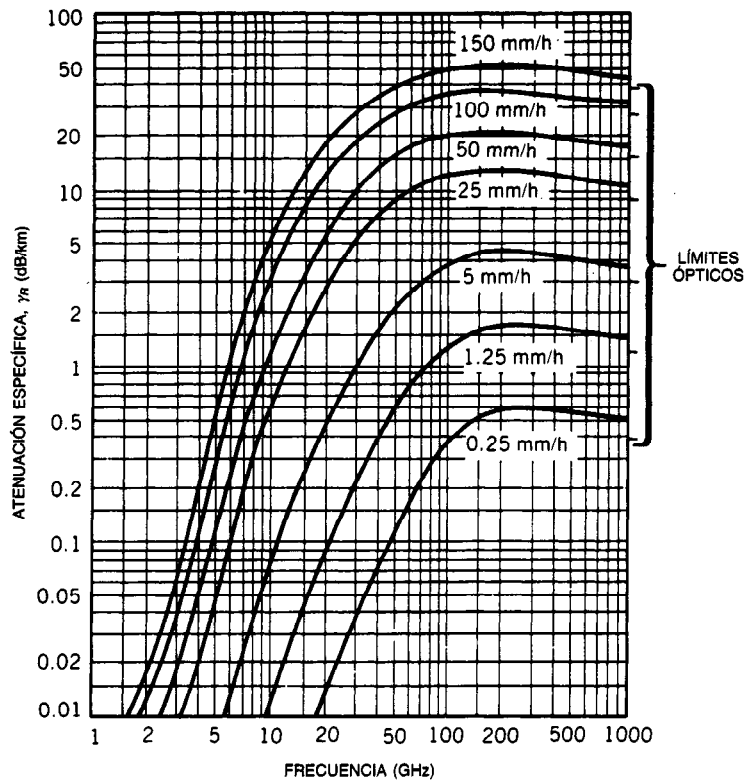


Fig. 49-60. Atenuación específica  $\gamma_R$  debida a la lluvia. Atenuación total  $A = \gamma L_s$ , donde  $L_s$  = longitud de la trayectoria inclinada. De Recommendations and Reports de la CCIR<sup>28</sup>, informe 721-1, con permiso.

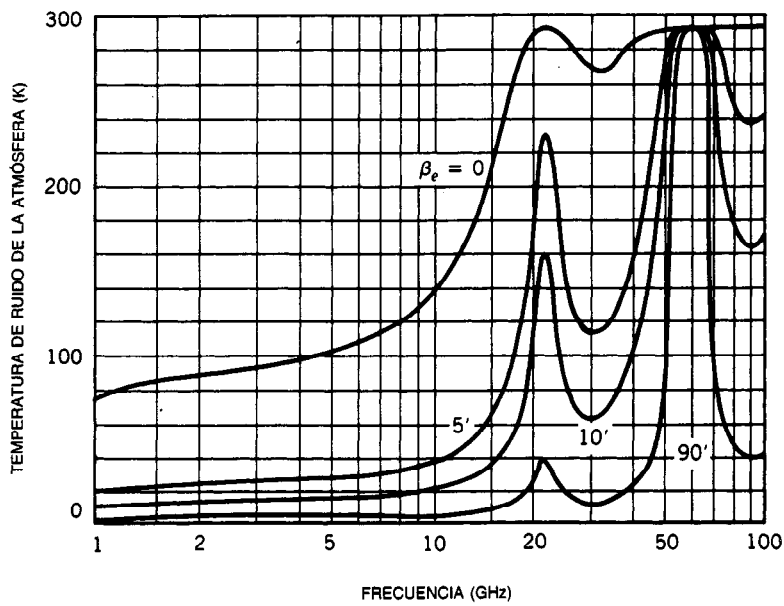


Fig. 49-61. Temperatura de ruido del cielo despejado (K).  $\beta_e$  = ángulo de elevación. Tomado de Miya<sup>10</sup>, con permiso.

donde cuatro haces de satélite están realmente presentes. La zona "occidental" se encuentra irradiada por dos haces polarizados ortogonalmente,  $A_V$  y  $A_H$ . De forma parecida, los haces polarizados ortogonalmente  $B_V$  y  $B_H$  irradian la zona "oriental". Los dos haces occidentales deben contar con lóbulos laterales bajos en todas partes dentro de la zona de cobertura oriental, y viceversa. Asimismo, ambos deben mantener un alto grado de pureza de polarización, y las estaciones terrestres en cada zona deben mantener la ortogonalidad de la polarización. El aislamiento del haz entre cualesquiera dos haces siempre se mide en la localización receptora y está dado por la razón de la potencia recibida de los dos haces. Estos niveles de potencia, a su vez, dependen de los niveles de potencia transmitida, de las ganancias de la antena en la *dirección de la antena receptora* y del factor de acoplamiento de polarización entre las antenas. Para una onda con razón axial de voltaje  $r_W$  y ángulo de inclinación  $\tau_W$  y una antena receptora con razón axial de voltaje  $r_A$  ( $0 \leq r_A$ ,  $r_W \leq 1$ ) y ángulo de inclinación  $\tau_A$ , el factor de acoplamiento de polarización  $\gamma$  de la ecuación 49-15 es:

$$\gamma = \frac{(1+r_W^2)(1+r_A^2) \pm 4r_W r_A + (1-r_W^2)(1-r_A^2)\cos 2\delta}{2(1+r_W^2)(1+r_A^2)} \quad (49-23)$$

donde  $\delta = \tau_W - \tau_A$  es la diferencia de los ángulos de inclinación y el signo  $(-)$  se utiliza para sentido opuesto de polarización. En el INTELSAT suelen utilizarse polarizaciones circulares duales ( $r \approx 1$ ); en muchos sistemas domésticos se utilizan polarizaciones lineales duales ( $r \approx 0$ ).

Los valores típicos para el aislamiento del haz entre dos haces cualesquiera se encuentran en el intervalo de 27 a 33 dB. Por supuesto, la potencia total de interferencia es la suma de todas las potencias de interferencia<sup>34</sup>. Para el ejemplo de la figura 49-65 existen tres elementos de interferencia. La interferencia se parece al ruido y debe agregarse sobre una base de potencia a los niveles de ruido térmico y por intermodulación, reduciendo aún más la razón de portadora entre densidad de ruido de la ecuación 49-20. Obsérvese que, como el aislamiento se mide en la localización receptora, el aislamiento del haz del enlace ascendente se mide en el satélite.

Las implicaciones relacionadas con los patrones de las antenas del satélite son ligeramente distintas para los aislamientos de los enlaces ascendente y descendente. Mientras que los valores de 27 a 30 dB con condiciones climatológicas despejadas son típicos para niveles de polarización cruzada en la antena de la estación terrestre y el satélite con polarización dual, los

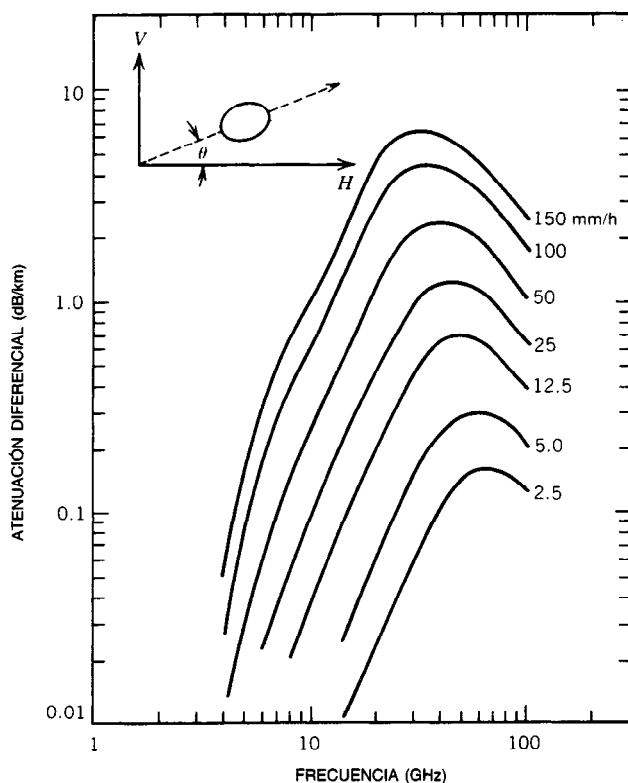


Fig. 49-62. Atenuación diferencial inducida por lluvia entre las polarizaciones I y II para diferentes índices de precipitación pluvial. Tomado de Hogg y Chu<sup>30</sup>, con permiso.

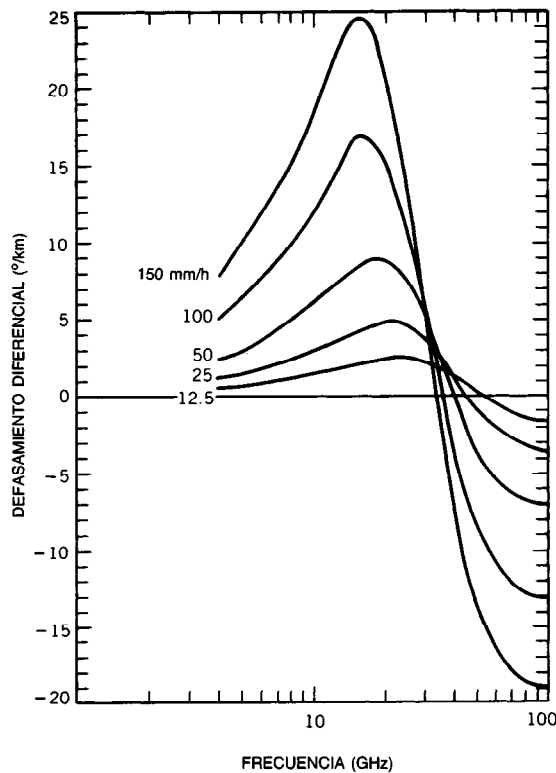


Fig. 49-63. Defasamiento diferencial inducido por lluvia entre las polarizaciones I y II para diferentes índices de precipitación pluvial. Tomado de Hogg y Chu<sup>30</sup>, con permiso.

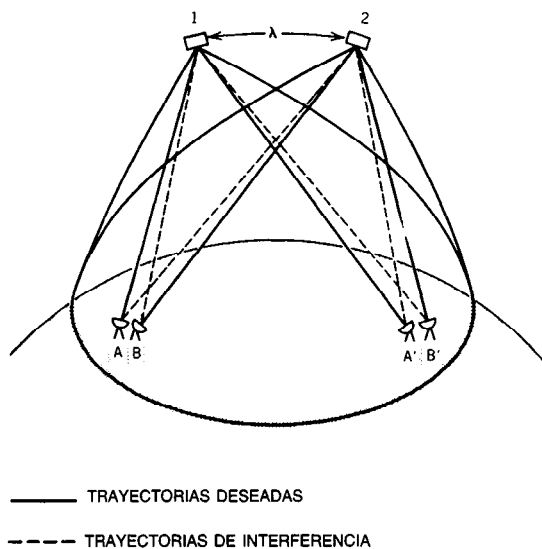


Fig. 49-64. Configuración geométrica de las trayectorias de interferencia con dos satélites. Ambos satélites cubren todas las estaciones.

efectos previamente observados de la despolarización por lluvia pueden degradar seriamente el aislamiento. Esto ha motivado el estudio de redes autoadaptables para compensación de la polarización<sup>35</sup>. La despolarización debida a la lluvia es peor para la polarización circular y para la polarización lineal en el caso de gotas de lluvia con ángulo de inclinación de  $45^\circ$  respecto a la onda. La degradación es mucho menos grave para la polarización lineal si las componentes de la onda están orientadas en forma paralela y perpendicular con respecto a la vertical y horizontal locales en la Tierra<sup>30</sup>. Vistas desde el satélite, las líneas de la "vertical local en la Tierra" son radiales a partir del punto subsatélite, y los lugares geométricos de la horizontal local constante son círculos concéntricos. El ángulo  $\phi$ , en las figuras 49-51 y 49-54 define un ángulo de inclinación de polarización del satélite con respecto a  $x$ , que estará alineado con el vector local de la gravedad en la localización terrestre  $p$ .

#### 49.4.4 Sistemas de satélite representativos

El continuo y rápido crecimiento de los sistemas de satélite está tipificado por los satélites INTELSAT, cuyas características principales se muestran en la figura 49-66<sup>18,33,36,37</sup>. El INTELSAT I (Pájaro madrugador), lanzado en 1965, tenía dos transpondedores de 25 MHz que operaban a 6/4 GHz, y las celdas solares en el tambor giratorio producían sólo 40 W de potencia. Cada generación subsecuente representó un crecimiento sustancial en cuanto a potencia y capacidad de canales. El INTELSAT IV<sup>33</sup> tenía doce transpondedores de 36 MHz, que utilizaban por completo la banda C disponible de 500 MHz. El satélite INTELSAT IV-A<sup>36-38</sup> fue el primero en emplear la reutilización de frecuencia por medio de haces espacialmente aislados, perfilados para cubrir los hemisferios occidental y oriental, según se observan desde la órbita. Esto permitió un incremento del 50 % en la capacidad de canales. El INTELSAT V<sup>39</sup>, el primero estabilizado por el cuerpo, reutilizó aún más las bandas de 6/4 GHz mediante la superposición de un haz de "zona" polarizado ortogonalmente en la parte sur de cada haz de hemisferio. En la figura 49-67 se pueden observar los componentes primordiales del INTELSAT V. El INTELSAT VI<sup>40</sup>, que se muestra en la figura 49-68, contiene 48 transpondedores que operan en las bandas C y  $K_u$  y pueden utilizar el TDMA conmutado por satélite (SS-TDMA)<sup>27,33</sup> para producir una capacidad de más de 30 000 circuitos de habla.

Como ya se mencionó, en muchos sistemas se utiliza más de un satélite en una zona específica. En el sistema INTELSAT se utilizan varios satélites en la zona del Atlántico, situados sobre el arco orbital desde aproximadamente los  $325^\circ$  hasta los  $340^\circ$  de longitud este. El satélite que interconecta a todos los usuarios se denomina satélite primario. Un segundo satélite, que interconecta los sitios de tráfico intenso que deben instalar una segunda antena en la estación terrestre, se denomina satélite de trayectoria principal. Los sistemas de los satélites primario y de trayectoria principal son independientes.

En ciertas situaciones se demanda la interconexión de los sistemas. Actualmente esto se realiza en las estaciones terrestres, pero están siendo considerados los enlaces entre satélites<sup>18,41</sup> a fin de mejorar la sensibilidad. Cada satélite extra de trayectoria principal requiere un conjunto de antenas en la estación terrestre.

La cobertura del océano Atlántico por el INTEL-SAT VI que se muestra en la figura 49-69 permite una reutilización por partida séxtuple de las bandas de 6 y 4 GHz en el satélite, así como haces puntuales en la banda  $K_u$ . Existen cuatro haces perfilados copolarizados para cobertura de zona espacialmente aislados (el perfilado de haces se describe en la sección 49.4.5) superpuestos en dos haces hemisféricos que están copolarizados y espacialmente aislados entre sí, pero polarizados ortogonalmente con respecto a los haces de la zona. Las polarizaciones circulares se utilizan con las zonas del enlace descendente en polarización sinistrorsa circular (LHCP, de *left-hand circularly polarized*) y los haces hemisféricos del enlace descendente en polarización dextrorsa circular (RHCP, de *right-hand circularly polarized*). Los haces de enlaces ascendentes poseen sentidos opuestos a los de sus contrapartes de enlaces descendentes. Los haces puntuales en la banda  $K_u$  están polarizados linealmente. Los aislamientos entre haces se especifican de modo que sean de 27 a 30 dB. Las características de los transpondedores son funciones de cada haz específico e incluyen (G/T) de  $-2$  a  $-14$  dB/K, densidades de flujo de saturación de  $-67$  a  $-87$  dBW/m<sup>2</sup>, dependiendo de las posiciones de los atenuadores de ganancia por pasos (como en la fig. 49-47). Las EIRP de enlaces descendentes varían de 23.5 a 44.4 dBW. En la figura 49-70 se muestra el plan de frecuencias del transpondedor. La duración prevista (vida útil de diseño) del satélite es de siete años; se intenta que el mantenimiento de posición de la estación sea  $\pm 0.1^\circ$  para longitud y latitud, y se espera que el subsistema para el control de orientación la mantenga apuntando a  $\pm 0.10^\circ$  en balanceo y cabeceo y a  $0.11^\circ$  en guiñada. El panel solar de 2 kW se complementa por baterías de níquel-hidrógeno con más de 2.2 kWh de capacidad.

En la tabla 49-14 se enumeran los parámetros para algunos<sup>42-44</sup> de los muchos sistemas que se utilizan alrededor del mundo. En la figura 49-71 se muestran los contornos EIRP de cobertura por haces perfilados del satélite GSTAR<sup>42</sup> para estados contiguos de la Unión Americana (CONUS, de *contiguous United States*) y Hawaii. Este satélite es representativo de la tendencia hacia el empleo de banda  $K_u$ , haces perfilados y reutilización de frecuencia por medio de polarizaciones lineales duales. Las formas del patrón (perfilado) se logran con bocinas de alimentación múltiple que se proyectan descentradas sobre los paraboloïdes. El diagrama de bloques del transpondedor del GSTAR (fig. 49-72) ilustra la canalización, la redundancia del receptor y las redes de formación de haces de antena. A fin de obtener distintas formas de cobertura discreta para varios canales se utilizan divisores de potencia variables; por ejemplo, para la cobertura de sólo la mitad de CONUS. Las "redes de modo dual", que se analizarán más tarde, permiten la combinación en una antena de canales contiguos del transpondedor, mientras que los multicanalizadores solamente requieren dar cabida a canales alternos (p. ej., números pares e impares de canales), lo cual simplifica su diseño.

En cuanto los sistemas se desplacen hacia la banda  $K_u$  a medida que se congestione la banda C, las tendencias futuras serán hacia frecuencias aún mayores, por ejemplo, las bandas de 30/20 GHz, y el empleo de haces puntuales para mayor capacidad continuará su desarrollo<sup>45</sup>. En los sistemas de satélites militares se utilizan haces múltiples como protección contra la perturbación (interferencia) intencionada<sup>46</sup>, y también existe una tendencia hacia frecuencias superiores<sup>47</sup>. Otros sistemas de satélites, tales como el INMAR-SAT<sup>48,49</sup>, atienden necesidades especiales, como las comunicaciones barco-costa en la banda L con usuarios móviles. El satélite para seguimiento y retransmisión de datos (TDRSS, de *tracking and data relay satellite*)<sup>50</sup> retransmite datos entre satélites en órbitas bajas y la Tierra a través de la órbita geoestacionaria. Los satélites de transmisión directa permiten la transmisión de TV, datos y otros servicios directamente al hogar<sup>51</sup>.

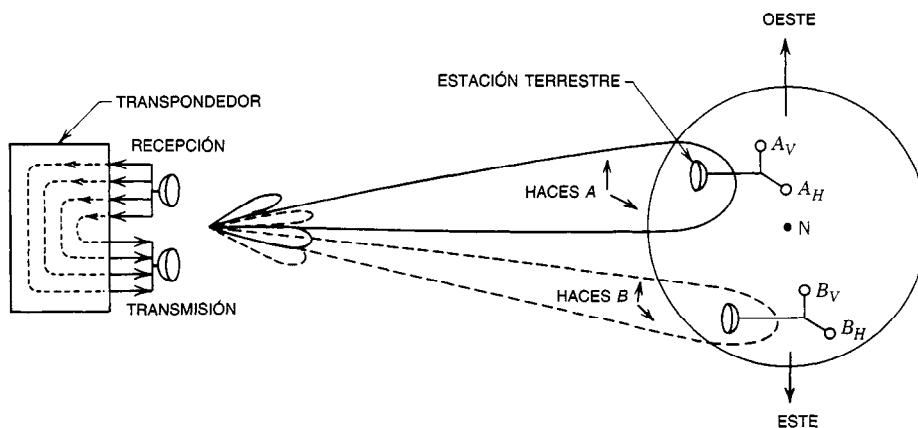
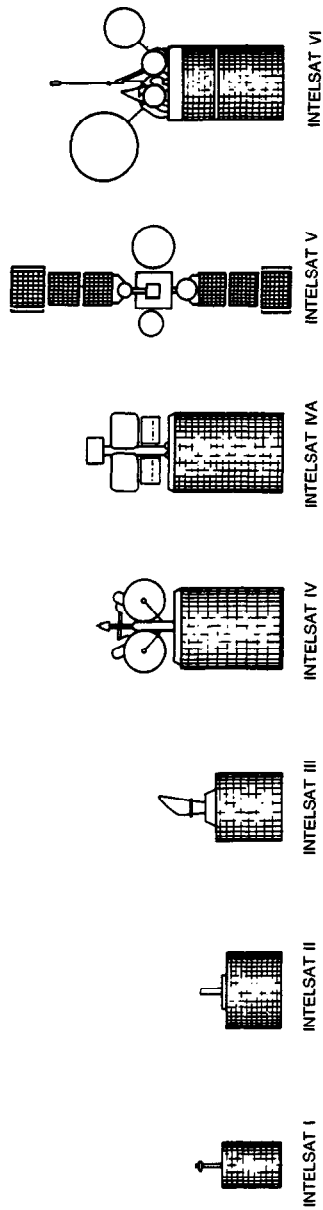


Fig. 49-65. Reutilización de frecuencias por medio de polarización dual y separación espacial del haz para cuatro haces de enlace ascendente y descendente del satélite.





	INTELSAT I	INTELSAT II	INTELSAT III	INTELSAT IV	INTELSAT IVA	INTELSAT V	INTELSAT VI
Primer lanzamiento	1965	1967	1968	1971	1975	1980	1986
Dimensiones (cm)							
Diámetro	72.1	142	142	238	238	1 560 (envergadura)	360 (tambor)
Altura	59.6	67.3	104	282 (tambor)	282 (tambor)	640	1 163 (global)
Masa (kg)							
En el momento del lanzamiento	68	162	293	1 385	1 469	1 928	3 740
En órbita	38	86	152	700	790	1 037	2 225
Vehículo de lanzamiento		Mejorada	De gran tanque	Atlas	Atlas	Atlas	Ariane-ST
Potencia primaria (W)	Thor-Delta	Thor-Delta	Thor-Delta	Centaur	Centaur	Centaur	Centaur
Transpondedores	40	75	120	400	500	1 200	2 100
Ancho de banda por transpondedor (MHz)	2	1	2	12	20	27	48
Cobertura	25	130	225	36	32, 36	40, 80 240	40, 80, 160
Hemisferio norte		Global	Global	Global, haces puntuales	Global, haces puntuales	Global, regional, haces puntuales	Global, regional, haces puntuales
EIRP (dBW)	11.5	15.5	23	22.5 (global)	22 (global)	22.29 (4 GHz)	23.31 (4 GHz)
Número de circuitos telefónicos	240 (sin acceso múltiple)	240	1 200	33.7 (haz puntual)	29 (haz puntual)	44 (11 GHz)	41.44.4 (11 GHz)
Duración (años)	1.5	3	5	4.000 (promedio)	6 000 (promedio)	12 000	30 000
Costo anual de los circuitos (\$K)	30	10	2	1	1	7 +	10
						0.9	0.5

Fig. 49-66. Satélites INTELSAT (arriba) y sus especificaciones.

Tabla 49-14. Propiedades de algunos sistemas de satélite

	GSTAR <sup>42</sup>		ANIK-C <sup>43</sup>		ARABSAT <sup>44</sup>	
Fecha de lanzamiento	1984	1982, 1984	1982, 1984	1984		
Vehículo de lanzamiento	STS <sup>a</sup> , Ariane 3 o Delta 3920	STS o Delta 3910	STS o Delta 3910	Ariane, STS		
Propietario	GTE Corp., USA	TELESAT, Canadá	TELESAT, Canadá	ARABSAT		
Aplicaciones	Digitales de 60/90 Mb/s, habla, imagen	Mensajes, TV, digitales a 91 Mb/s	Mensajes, TV, digitales a 91 Mb/s	Comunicaciones, TV para la comunidad a 2.5 GHz		
Bandas de frecuencia, ascendente/descendente (GHz)	14-14.5/11.7-12.2	14-14.5/11.7-12.2	14-14.5/11.7-12.2	5.925-6.425/3.7-4.2; 2.5		
Cobertura	CONUS, Alaska, Hawai	Sur de Canadá	Sur de Canadá	Norte de África y Medio Oriente		
Número de transpondedores	16 (8 en cada polarización)	16	16	25 a 6/4, 1 a 2.5 GHz		
Ancho de banda de los transpondedores (MHz)	54	54	54	33		
Polarizaciones	Lineal dual, aislamiento = 33 dB	Lineal dual	Lineal dual	Polarización circular dual		
EIRP (dBW)	42 (CONUS)	47	47	31 a 4 GHz; 39 a 2.5 GHz		
Potencia de salida del amplificador	5 a 20 W, 2 a 30 W	15	15	8.5 a 4 GHz; 50 W a 2.5 GHz		
Fabricante	RCA	Hughes Aircraft	Hughes Aircraft	Aerospatiale/Ford		
Estabilización	Por el cuerpo	Por giro	Por giro	Por el cuerpo		
Masa en órbita (kg)	1250	567	567	1195		
Potencia del panel solar (W)	1900	900	900	1285		

<sup>a</sup> STS, sistema de transporte espacial

## 49.4.5 Subsistemas

## Subsistemas de satélites

Los satélites para comunicaciones deben funcionar confiablemente en el ambiente espacial, por lo común durante periodos de 7 a 10 años. Como consecuencia de este requisito, en el diseño del vehículo espacial deben tomarse en cuenta factores tales como el vacío [especialmente el vacío parcial ( $\approx 10^{-6}$  torr) en la órbita de transferencia, donde podría ocurrir descarga disruptiva]; ambiente térmico, que provoca extremos de temperatura hasta  $200^{\circ}\text{C}$ ; radiación, incluyendo electrones y protones, que podría dañar los semiconductores y que degrada el rendimiento de las celdas solares; radiación ultravioleta, que puede dañar algunos materiales; meteoroides; y el campo geomagnético<sup>10,52</sup>.

Los subsistemas más importantes para un satélite se enumeran en la tabla 49-15, en la que se indican, para

el satélite INTELSAT VI, el porcentaje de masa seca total y el porcentaje de potencia primaria requeridos por cada subsistema.

Los amplificadores de salida del transpondedor consumen la mayor parte de la potencia, y el subsistema de comunicaciones, que abarca las antenas y dispositivos electrónicos del transpondedor, es el que cuenta con un peso más elevado.

## Antenas

Las antenas de los satélites han evolucionado hasta llegar a ser un factor importante en el diseño y peso de cada sistema. Esto es esencialmente el resultado de incorporar la reutilización de frecuencia y el perfilado de haces. Este último se ilustra en el diagrama simplificado de la figura 49-73. Cada alimentación desplazada del foco,  $f$ , del paraboloide descentrado produce un "haz componente" de exploración con ancho de banda de 3 dB,  $\theta_c \approx 65 \lambda/D$ , donde  $D$  es la abertura

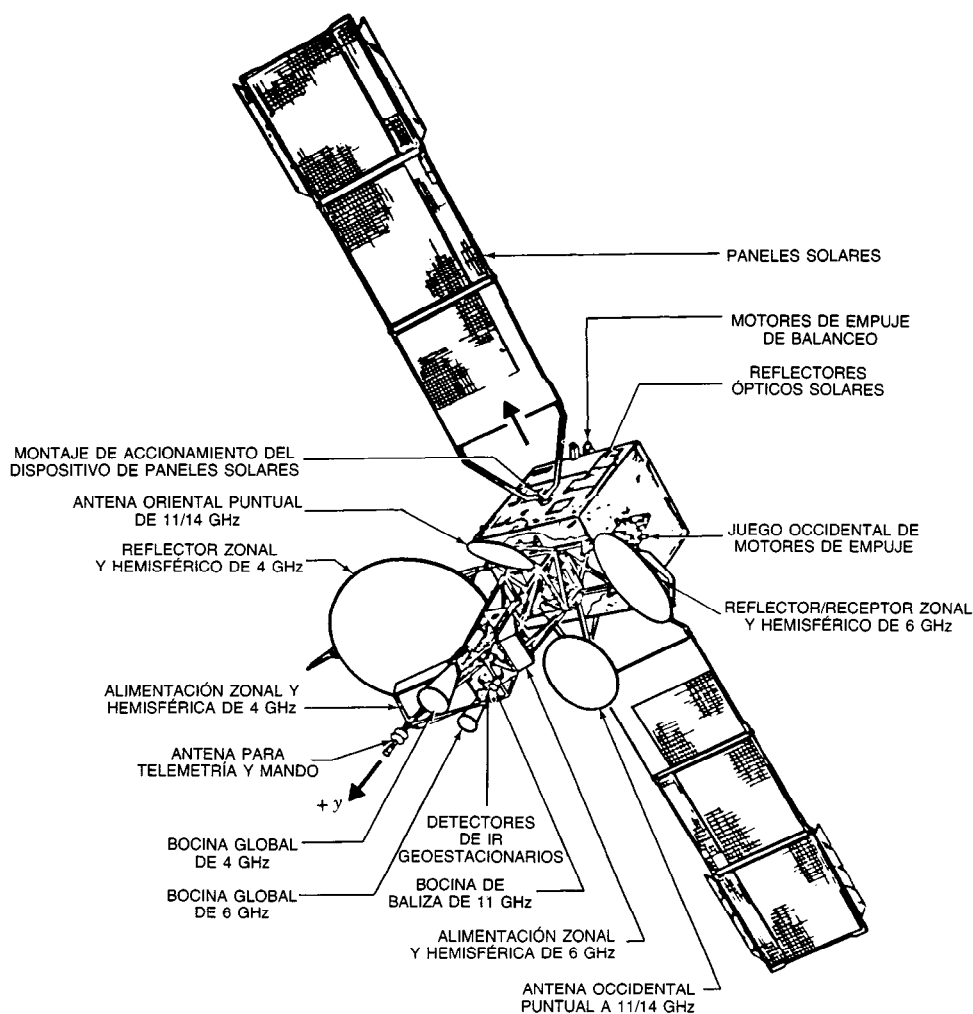


Fig. 49-67. El satélite INTELSAT V tiene una longitud total de más de 15 m. Cortesía de la Ford Aerospace and Communications Corp.

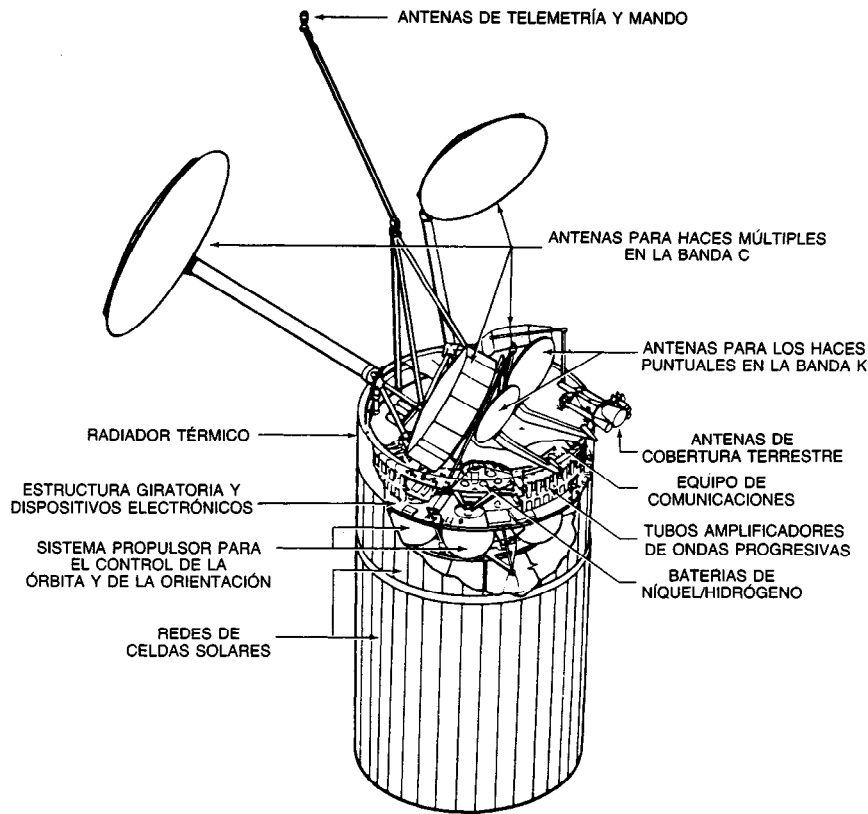


Fig. 49-68. Satélite INTELSAT VI. Cortesía de la Hughes Aircraft Co.

proyectada del diámetro. Se utilizan antenas de alimentación desplazada, porque se encuentran libres de bloqueo, lo que podría degradar los lóbulos laterales. Una red de formación de haces, tal como un divisor de potencia, suma coherentemente los haces a fin de formar un haz compuesto, como se muestra en el corte transversal y en la base de un contorno en la figura. Con muchos sistemas de alimentación es posible construir las deseadas configuraciones de haces a fin de seguir los contornos de cobertura deseados, tales como la frontera de un país. Para una mayor abertura, los haces componentes son más pequeños, y se requieren más para llenar la cobertura, con lo que se incrementa la complejidad del equipo, pero también permite un mejor control del patrón.

La mínima ganancia de cobertura está relacionada de forma aproximada con el área del contorno del haz alrededor del contorno de  $-4$  a  $-85$  dB,  $\Omega$  en grados al cuadrado, como:

$$g = \frac{k}{\Omega}$$

donde  $k$  es típicamente igual a 13 000, pero puede variar de 10 000 a 15 000, dependiendo del diseño y del diámetro del reflector. Para dos haces perfilados que deben estar espacialmente aislados, una regla útil es

que la separación de borde a borde,  $\sigma$  grados, que mantiene un aislamiento entre haces de 27 dB entre contornos adyacentes con cobertura de  $-4$  dB, debe ser de por lo menos  $1.5 \theta_c$ . Esto implica que:

$$\sigma \approx \frac{100}{(D/\lambda)} \text{ grados} \quad (49-24)$$

Para un sistema con separación específica entre haces puntuales adyacentes, el diámetro mínimo de la antena está determinado por la ecuación 49-24. Este resultado supone que los haces no están sustancialmente degradados por aberraciones de barrido fuera de eje, lo que a su vez implica reflectores o lentes con  $f/D \geq 1$ . La determinación de las excitaciones de alimentación óptimas en amplitud y fase a fin de producir de forma simultánea elevada ganancia de cobertura y bajos lóbulos laterales es un problema importante del diseño. Las coberturas de los haces se optimizan por medio de técnicas auxiliadas por computadora. El control de la polarización requiere avances que presenten una buena simetría de patrones para la polarización circular. A fin de lograr polarizaciones lineales puras en los reflectores descentrados, la superficie de éstos puede estar compuesta por rejillas paralelas o, alternativamente, es posible utilizar rejillas de difracción en el plano de abertura. Durante la década de

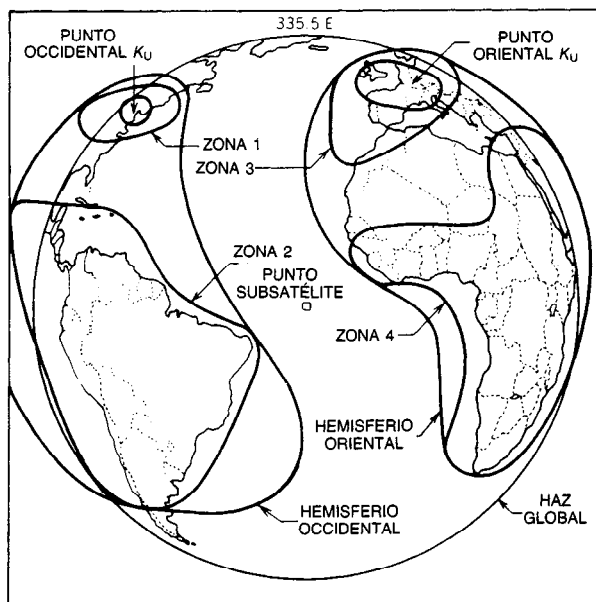
1970 y a principios de la década de 1980 los diseños de antenas para satélites recibieron mucha atención<sup>34,53</sup>. Si la red de la figura 49-73 tiene dos puertas de entrada y dos puertas de salida y es una red ortogonal (es decir, sin pérdidas) de puertas múltiples, entonces cada puerta de entrada produce haces perfilados nominalmente idénticos, donde los haces componentes pueden considerarse como sumados en cuadratura. Las dos puertas de entrada están aisladas siempre que las alimentaciones estén acopladas. Esta red en "modo dual" puede generalizarse a fin de contar con muchas puertas de alimentación, y es de utilidad en los transpondedores de satélite en los cuales la separación de frecuencia entre canales adyacentes es muy pequeña y sería difícil diseñar multicanalizadores contiguos. En tales transpondedores es posible utilizar dos conjuntos de multicanalizadores, cada multicanalizador con canales alternativos; por ejemplo, canales con numeración par e impar. La conexión de cada multicanalizador a una de las puertas de la red produce haces casi idénticos para cada "modo", y la red ortogonal mantiene el aislamiento entre las puertas de entrada a fin de evitar la interferencia de canales adyacentes.

### Transpondedores

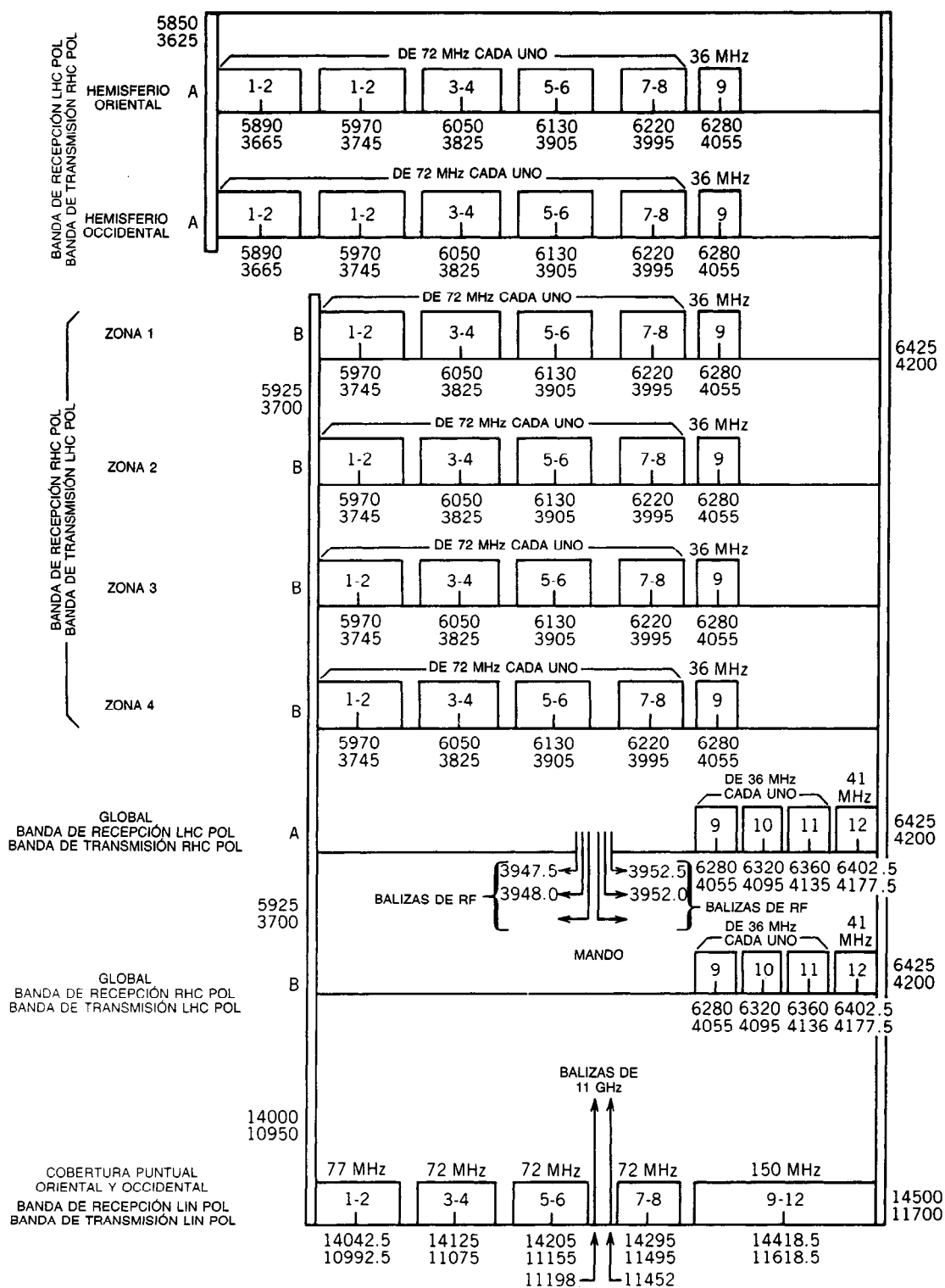
El subsistema del transpondedor incluye el receptor, filtros, conmutadores de redundancia, multicanalizador de entrada, amplificadores y atenuadores, amplificador de salida, multicanalizador de salida, y todos los conmutadores para las interconexiones idóneas de haces de enlaces ascendente y descendente (fig. 49-72). En la referencia 36 se describen los componentes del

transpondedor. En el diagrama de un receptor de conversión única típico, como el que se muestra en la figura 49-74<sup>41</sup>, podría utilizarse un amplificador a diodo túnel como amplificador de bajo ruido (LNA, de *low-noise amplifier*) de banda C, pero los transistores de efecto de campo (FET, de *field effect transistor*) de GaAs están adquiriendo importancia, especialmente a altas frecuencias<sup>36</sup>. La ganancia del LNA debe ser suficientemente alta, en general por lo menos de 10 a 15 dB, a fin de establecer la temperatura de ruido del sistema. Una ganancia global típica del receptor es de aproximadamente 70 dB, y la ganancia global típica de un transpondedor de satélite, incluyendo el amplificador de salida, se encuentra en el intervalo de 100 a 130 dB. Dado que la antena del satélite suele estar a unos 300 K debido a su orientación hacia la Tierra, existe un límite sobre la mejora en G/T obtenible mediante la reducción del factor de ruido del LNA del satélite. Por el contrario, una estación terrestre "ve" casi siempre cielo frío, por lo que la temperatura de su antena es menor y su LNA debe tener de forma correspondiente una temperatura de ruido menor.

Los filtros de RF deben limitar el ancho de banda del ruido del sistema sin distorsionar el espectro de la señal portadora debido a la respuesta de no linealidad de la amplitud y al retardo de grupo. La distorsión por retardo de grupo puede compensarse algunas veces incluyendo un ecualizador en la trayectoria de la señal. La degradación de filtro para las portadoras digitales puede dar por resultado una interferencia entre símbolos<sup>41</sup>. Otros parámetros que es necesario controlar son las salidas espurias del oscilador local (LO, de *local*



**Fig. 49-69.** Contornos de la cobertura de la antena del INTELSAT VI para la zona del océano Atlántico (AOR, de *Atlantic Ocean Region*) desde la localización del satélite primario. Los haces zonales, hemisféricos y globales están en los 6/4 GHz, y los haces puntuales oriental y occidental están en los 14/11 GHz.



**Fig. 49-70.** Frecuencias centrales, anchos de banda y polarización de los transpondedores del INTELSAT VI. LHC POL = con polarización sinistrorsa circular; RHC POL = con polarización dextrorsa circular; LIN POL = con polarización lineal. Tomado de Ghais y colaboradores<sup>20</sup>, con permiso.

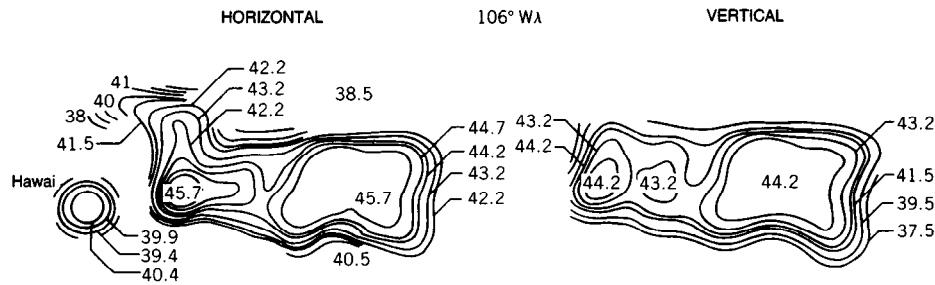


Fig. 49-71. Contornos EIRP GSTAR para estados contiguos de la Unión Americana (CONUS) y Hawai. Tomado de Lovell y Cuccia<sup>45</sup>, con permiso.

oscilator), estabilidad de frecuencia del LO, salidas espurias y trayectorias múltiples de los canales adyacentes en los multicanalizadores, productos de intermodulación en los amplificadores de salida, y para portadoras digitales, respuestas transitorias de pulsos.

Típicamente para el amplificador de entrada se utilizan TWT, aunque también se están empleando amplificadores de estado sólido en la banda C. Los TWT están disponibles para muchas de las bandas de frecuencia de interés en las comunicaciones por satélite, y pueden proporcionar un gran intervalo de niveles de potencia de salida. El rendimiento típico de TWT de banda C incluye<sup>54</sup> las eficiencias (DC a RF) del 35 al 45%, potencia de salida de 1 a 15 W, anchos de banda de 1 a 700 MHz, y ganancias de 30 a 55 dB. También existen mayores ganancias y TWT con mayor potencia de salida a frecuencias superiores. Los TWT típicos para satélites en la banda K<sub>u</sub> operan entre 20 y 50 W, y en un satélite para transmisión directa<sup>50</sup> se utiliza un TWT de 200 W a entre 12.25 y 12.75 GHz. Las degradaciones en el rendimiento del sistema debidas a los TWT surgen por su respuesta no lineal de amplitud y su defasamiento cerca de la saturación. Algunas dificultades incluyen la conversión de AM a PM, indicada por el coeficiente de conversión AM/PM del tubo, que por lo general varía de 2 a 7°/dB, y el defasamiento del tubo, que puede variar de 15 a 50° en saturación<sup>16</sup>.

Los multicanalizadores de entrada y salida constituyen una parte importante del volumen y peso de los transpondedores. Los primeros diseños se basaban en las respuestas de Tchebycheff, pero la síntesis de los filtros de función elíptica de modo dual prometen una reducción de peso<sup>55,56</sup>. En la figura 49-75 se muestra un filtro de cavidad de modo dual en el que se generan dos cavidades eléctricas en el volumen de una cavidad física por medio de modos ortogonales de guías de ondas. Los tornillos de acoplamiento proporcionan acoplamientos importantes dentro de cada cavidad a fin de producir altos valores de  $Q$  ( $> 10\,000$ ) y la respuesta elíptica de la función. Los modos de orden superior pueden producir valores de  $Q$  mayores. Los filtros de resonador dieléctrico ofrecen una reducción de peso.

#### Otros subsistemas

Van Trees<sup>36(p. 451)</sup> describe los demás subsistemas de un satélite. Para las funciones de telemetría, seguimiento y mando (TT&C, de *telemetry, tracking and*

*command*), los datos de telemetría de los detectores y dispositivos a bordo del vehículo espacial para monitorizar su condición suelen muestrearse, multicanalizarse en tiempo y emplearse para modular en fase una portadora de enlace descendente a una frecuencia asignada. La velocidad de transferencia de bits es baja, por ejemplo 4 800 b/s para el INTELSAT VI, y es aún menor para vehículos espaciales menos complicados. Las señales de mando FM de baja velocidad de transferencia de bits son recibidas en la banda de enlace ascendente y deben decodificarse de forma confiable. El enlace de comando debe ser altamente inmune al ruido, para prevenir errores y la posibilidad de falsos comandos. Los satélites complicados deben distinguir varios cientos de comandos. Varias antenas omnidireccionales, por lo general bicónicas, se utilizan para la órbita de transferencia, y en la generalidad de los casos suele emplearse una pequeña bocina para el funcionamiento en órbita. Un método de seguimiento consiste en modular la portadora del mando con tres tonos relacionados (p. ej., de 283.4, 3 968 y 27 777 Hz para el INTELSAT VI), permitiendo la determinación de una única distancia de alcance. En otra técnica se emplean pulsos, y el retardo medido indica el alcance.

El subsistema de energía eléctrica suministra potencia de las celdas solares y también proporciona regulación y acondicionamiento de los voltajes de la barra colectora. La energía solar disponible es aproximadamente 1.35 kW/m<sup>2</sup>, y las eficiencias típicas de las celdas son del 13%. En el periodo aproximado de siete años que dura el satélite, el flujo de electrones del ambiente espacial suele degradar la salida de las celdas en aproximadamente el 30 o el 40%<sup>16</sup>. Las baterías deben ser capaces de mantenerse en funcionamiento durante los eclipses, que pueden durar hasta 72 minutos diarios durante los equinoccios. Los vehículos espaciales estabilizados por el cuerpo pueden tener toda la superficie de los paneles solares dirigida permanentemente al Sol, pero un satélite estabilizado por giro sólo tiene aproximadamente la tercera parte de sus celdas expuestas en cualquier instante. Para ambos tipos de satélite, entre los equinoccios y los solsticios el ángulo solar varía debido a la inclinación de la eclíptica de 23.5°, por lo que se reduce aún más la potencia.

La temperatura en el espacio de una pequeña esfera negra de radio  $r$  en equilibrio térmico se determina igualando el flujo solar de  $1.35 \times 10^3$  W/m<sup>2</sup> con el calor

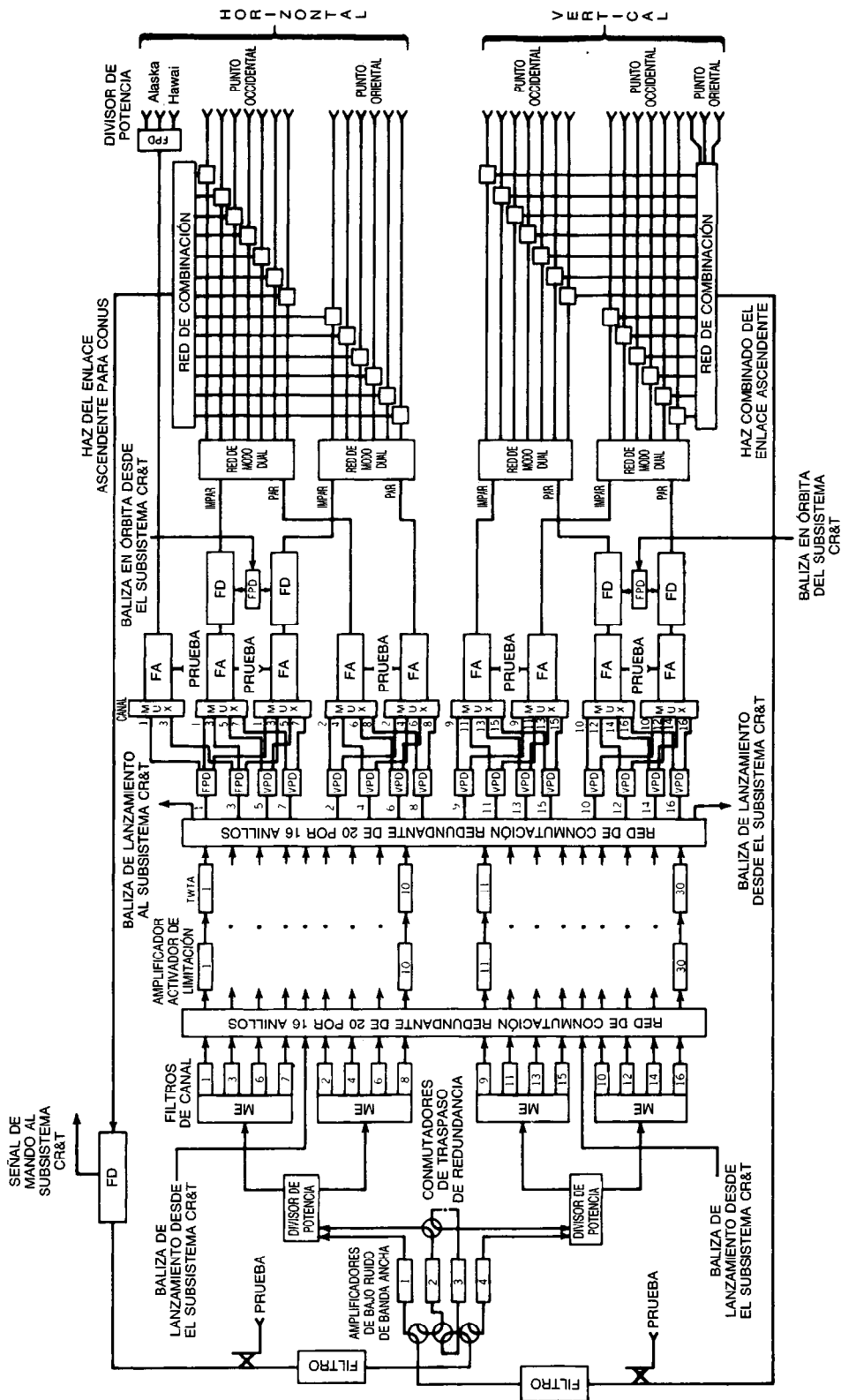


Fig. 49-72. Diagrama de bloques para el transpondedor del GSTAR. D = bicanalizador; FPD = divisor de potencia fija; VPD = divisor de potencia variable; MUX = multicanalizador; TWTA = amplificador de tubo de ondas progresivas; FD = filtro de dirección; ME = multicanalizador de entrada; FA = filtro de armónicos. Tomado de Napoli<sup>42</sup>, con permiso.



radiado por la esfera:  $4\pi^2 \sigma T^4$ , donde  $\sigma$ , la constante de Stefan-Boltzmann, es  $5.67 \times 10^{-8} \text{ W/m}^2/\text{K}^4$ . Se encuentra que la temperatura  $T$  es aproximadamente 280 K. En ausencia de cualquier entrada solar, la temperatura debería ser 4 K. En la práctica, la diferencia de temperatura entre la luz solar y la sombra puede ser de aproximadamente 200 K. La radiación y la conducción son los únicos mecanismos para la eliminación de calor. Los subsistemas de control térmico incluyen materiales reflejantes pasivos a fin de proteger los componentes, y posiblemente también cuenten con calentadores térmicos para mantener reducidas las variaciones de temperatura en algunos dispositivos electrónicos de los vehículos espaciales.

La confiabilidad de un componente o de un sistema, que determina la elección de los materiales, diseños y redundancias, es la probabilidad de que dicho componente o sistema sea funcional después de un tiempo prescrito. La confiabilidad del INTELSAT VI es de 0.996 durante los dos primeros meses de funcionamiento, y de 0.781 durante su tiempo de vida de siete años.

#### Subsistemas de estaciones terrestres<sup>10,17,57</sup>

Los principales elementos de una estación terrestre de comunicación por satélite son la antena, el receptor de bajo ruido (LNR, de *low noise receiver*)/convertidor descendente, el convertidor ascendente, el (los) amplificador(es) de alta potencia (HPA, de *high-power amplifier*), el equipo terrestre para la comunicación de superficie (GCE, de *ground communication equipment*) y el equipo de monitorización y control, como se muestra en la figura 49-76.

Existe una gran variedad de estaciones terrestres, que van desde las grandes estaciones del INTELSAT hasta pequeñas terminales a bordo de barcos.

En los sistemas de satélites para radiodifusión directa se utilizan terminales receptoras de tipo casero con antenas parabólicas de menos de 1 m que operan en la banda  $K_u$ . Por su parte, las antenas estándares A, B y C del INTELSAT poseen un intervalo de frecuencias y una relación G/T específicos. Para las antenas estándares A y B de 6/4 GHz, la relación G/T es:

$$\text{Estándar A: } G/T \geq 40.7 + 20 \log f/4 \quad \text{dB/K}$$

$$\text{Estándar B: } G/T \geq 31.7 + 20 \log f/4 \quad \text{dB/K}$$

Para la antena estándar C en la banda  $K_u$ , la especificación de G/T es más complicada debido al efecto del ruido atmosférico sobre la temperatura de la antena<sup>10(p. 341)</sup>. Las estaciones INTELSAT deben ecualizar la distorsión por retardo de grupo producida por el satélite. La ecualización lineal de retardo varía de  $\pm 10 \text{ ns/MHz}$  para un ancho de banda de 1.25 MHz a  $\pm 1 \text{ ns/MHz}$  para un ancho de banda de 36 MHz.

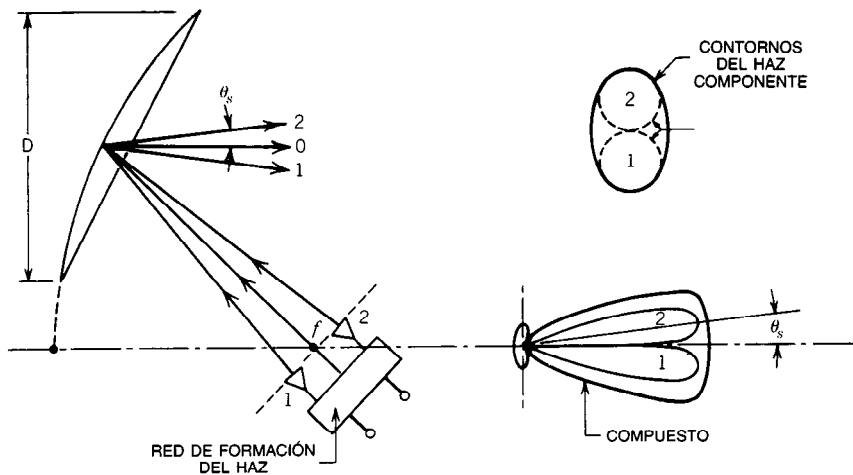


Fig. 49-73. Perfilado de haces con dos alimentaciones y un paraboloide descentrado.

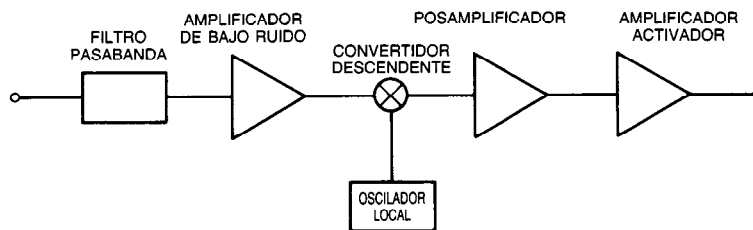


Fig. 49-74. Receptor de conversión sencilla.

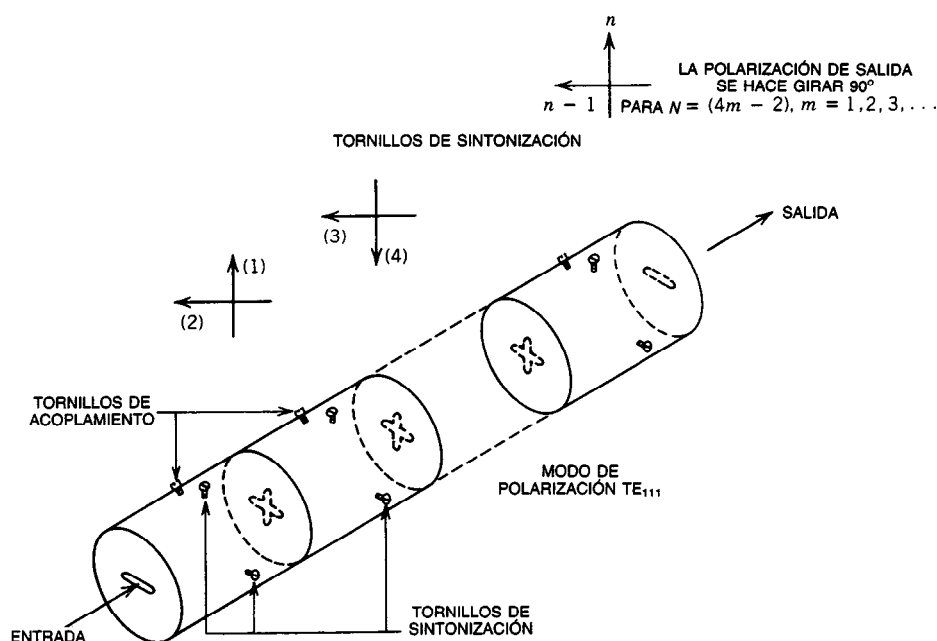


Fig. 49-75. Filtro longitudinal con cavidad de modo dual. Tomado de Atia y Williams<sup>55</sup>, con permiso.

La mayor parte de las antenas de las grandes estaciones terrestres son sistemas de Cassegrain, con las superficies reflectora principal y subreflector configuradas a fin de maximizar la eficiencia de apertura. Además, muchas grandes estaciones están construidas con sistemas de "guías de ondas del haz", que son una serie de espejos que guían la RF desde el subreflector hasta la bocina de alimentación. Por consiguiente, el sistema de alimentación y todo el equipo HPA pueden localizarse al nivel del piso, con lo que se minimizan las pérdidas debidas a corridas de guías de ondas entre el HPA y la alimentación. En antenas bien diseñadas se han logrado eficiencias de apertura hasta del 80%<sup>53</sup>. Varios sistemas de satélites deben tener acceso a más de un satélite a partir de un lugar. En las estaciones terrestres se han desarrollado antenas de haces múltiples, tales como las Torus<sup>57</sup>, a fin de lograr una visión hasta del 50% o el 60% del arco geoestacionario.

Los receptores de bajo ruido han sido amplificadores paramétricos enfriados criogénicamente en el sistema INTELSAT, pero ya están siendo sustituidos por amplificadores enfriados por efecto Peltier. Para sistemas más pequeños, se utilizan FET de GaAs cuyas temperaturas de ruido varían de 70 a 100 K.

El HPA, ya sea un TWT o un clistrón con potencia hasta de 10 kW, debe ser liberado sustancialmente de la saturación para operar con portadoras múltiples. Para la implementación del HPA se utilizan varios amplificadores pequeños y un multicanalizador de salida.

#### 49.4.6 Modulación y acceso

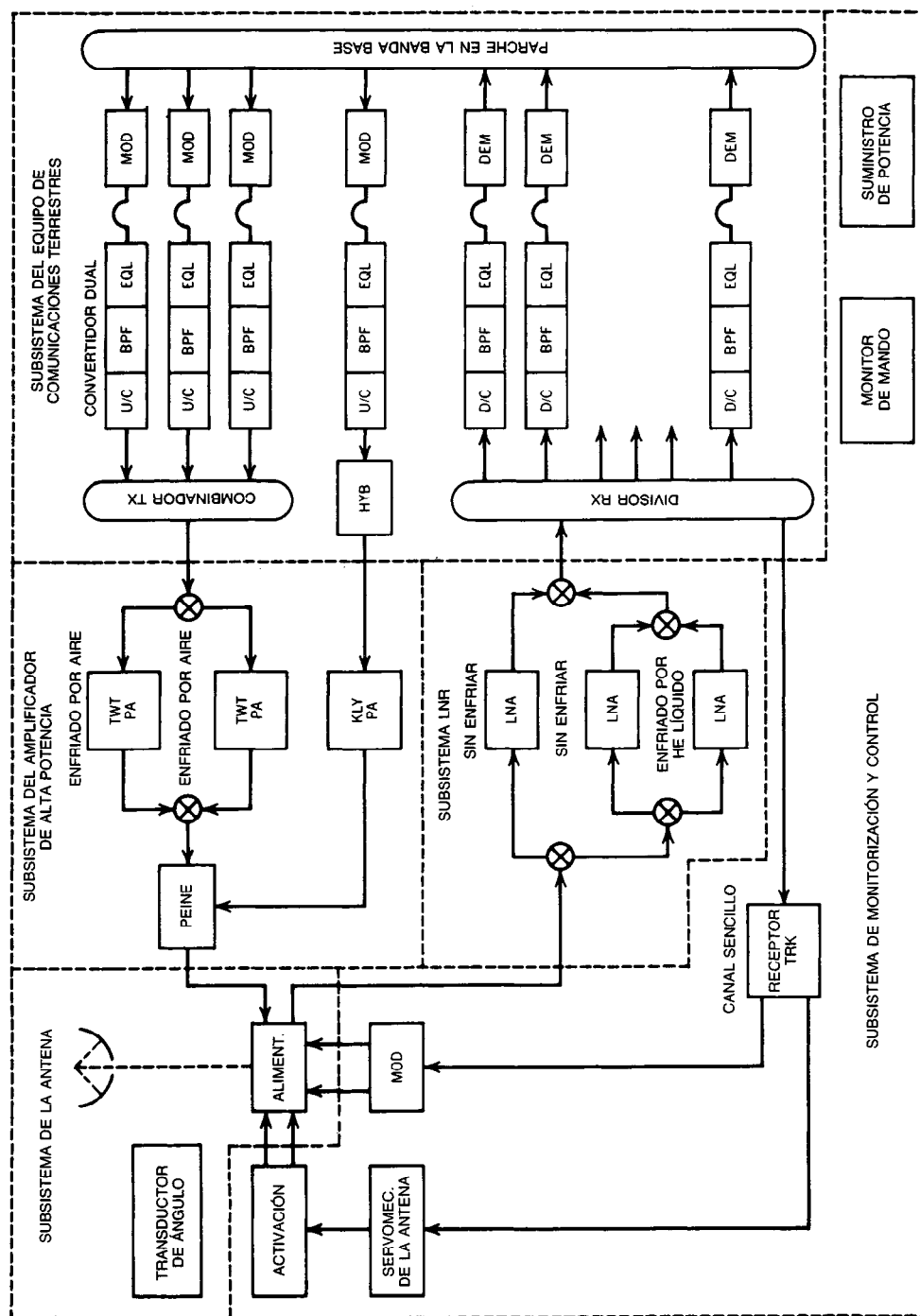
En las secciones 48.1 y 48.5 se describen las técnicas de modulación analógica y digital y las de acceso, por lo

que aquí solamente se describirán los puntos más destacados de aquellos aspectos relacionados con las comunicaciones por satélite; en la vasta literatura sobre el tema es posible encontrar mayores detalles<sup>10,16,17,19,26,27,36,41</sup>.

La modulación analógica, especialmente la FM, ha sido dominante en los sistemas de satélites. Sin embargo, la tendencia es claramente hacia la modulación di-

Tabla 49-15. Porcentajes del total de la masa y potencia de los subsistemas del INTELSAT VI

Subsistema	Masa (% de 1 570 kg)	Potencia (% de 2 000 W)
Antenas	18	0
Repetidores	19	74
Seguimiento, comando y telemetría	3	3
Orientación	5	1
Propulsión	8	1
Energía eléctrica	18	
Dispositivos electrónicos		1
Carga de baterías		13
Pérdidas		3
Incertidumbre de carga		2
Control térmico	4	2
Estructura	17	0
Integración	8	0



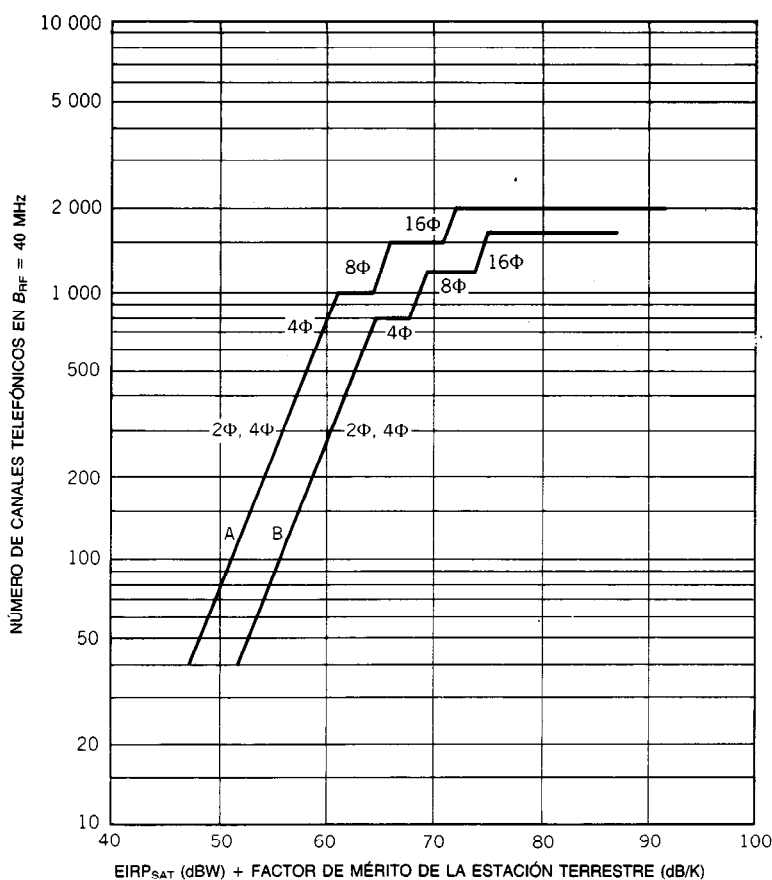
**Fig. 49-76.** Diagrama de bloques de una estación terrestre típica. HPA = amplificador de alta potencia; GCE = equipo terrestre de comunicaciones; TWT = tubo de ondas viajeras; PA = amplificador de potencia; KLY PA = amplificador de potencia clístron; LNA = amplificador de potencia clístron; LNR = receptor de bajo ruido; U/C = convertidor ascendente; BPF = filtro pasabanda; EQL = ecualizador; HYB = híbrido; MOD = modulador; DEM = desmodulador; D/C = convertidor descendente; SERVOMECH. = servomecanismo; ALIMENT. = alimentación. Tomado de E. R. Walthall, "Earth Station Technology", en Van Treecs<sup>6</sup>, págs. 580 a 587, con permiso.

gital. Asimismo, el acceso múltiple por división de frecuencia (FDMA, de *frequency division multiple access*) es con mucho el dominante, aunque ya se están utilizando sistemas de acceso múltiple por división de tiempo (TDMA, de *time division multiple access*).

Para el FDMA, a los usuarios se les asigna una única frecuencia portadora. Los "usuarios" son estaciones terrestres capaces de multicanalizar por división de frecuencia (FDM) muchos canales individuales de habla en una portadora común; por ejemplo, para el tráfico telefónico multicanalizado. Estas señales pueden modular la frecuencia de una portadora, y cada portadora define un acceso. Con la FM, este esquema se denomina FDM-FM-FDMA. La eficiencia de utilización del transpondedor, en canales de habla por transpondedor, decrece a medida que aumenta el número de accesos. Para una simple portadora, el TWT del satélite puede hacerse funcionar cerca de la saturación, maximizando la potencia para cada canal de habla. A medida que aumenta el número de portadoras en el transpondedor, cada una contiene menos canales de

habla. También el espectro de cada portadora es aproximadamente gaussiano, y debe haber alguna banda de guarda entre las portadoras. Además, el amplificador del satélite debe ser liberado de la saturación a fin de evitar los productos de intermodulación entre las portadoras. Un efecto común es reducir el número de canales de habla por transpondedor de 36 MHz, desde 1 000 para una sola portadora de banda amplia hasta aproximadamente 400 para 15 portadoras. Un transpondedor de 36 MHz por lo general portaría un canal de TV, pero en algunos sistemas hay dos canales por transpondedor.

La modulación digital ofrece la posibilidad de utilización mejorada. Es posible concebir varias combinaciones de multicanalización, modulación y acceso. Por ejemplo, una señal de la banda base puede modularse por codificación de pulsos (PCM, de *pulse code modulated*) y multicanalizarse por división de tiempo (TDM) con otras señales, que a su vez pueden modular por defasamiento (PSK, de *phase shift key*) una portadora de RF en un sistema TDMA. El esquema global



**Fig. 49-77.** Utilización del transpondedor para acceso múltiple mediante manipulación por defasamiento (PSK). Curvas A: TDM-PCM-CPSK-TDMA. Curvas B: TDM-PCM-CPSK-FDMA. TDM = multicanalización por división de tiempo; PCM = modulación por codificación de pulsos; CPSK = manipulación por defasamiento coherente; TDMA = acceso múltiple por división de tiempo; FDMA = acceso múltiple por división de frecuencia.

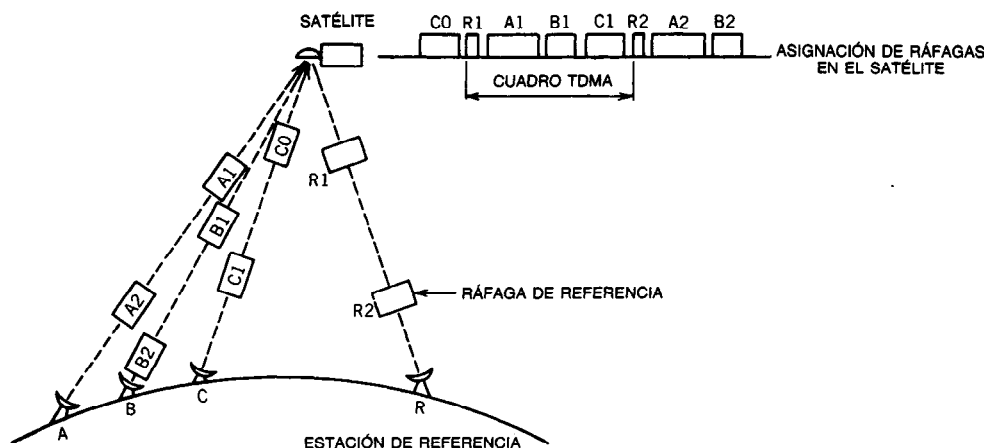


Fig. 49-78. Configuración básica de un sistema de acceso múltiple por división de tiempo (TDMA). Tomado de Miya<sup>10</sup>, con permiso.

se denominaría PCM-TDM-PSK-TDMA. Este esquema específico ofrece una alta utilización que no se degrada significativamente a medida que aumenta el número de acceso. En el método de un solo canal por portadora (SCPS, de *single-channel-per-carrier*) se asigna un canal de habla por portadora. Así se codifican típicamente, por ejemplo, 8 bits<sup>6</sup> a una velocidad de 8 kb/s para una velocidad de 64 kb/s por canal. El método SCPC permite una eficiencia de utilización constante. Para sistemas analógicos, la fidelidad de la señal es la medida del rendimiento y está relacionada con la razón de señal sobre ruido desmodulada. Para la FM, la relación de portadora de RF sobre ruido total típicamente debe ser mayor de 13 dB para una buena calidad de sonido. El criterio para las señales digitales es la probabilidad de error en la desmodulación de un bit. Para cada tipo de modulación digital, aquella está relacionada con la razón  $E_b/N_0$  (secciones 49.4.3 y 48.7). Para una probabilidad aceptable dada de error, el rendimiento, o eficiencia de utilización, puede ob-

tenerse para varias modulaciones digitales. En la figura 49-77 se ilustra la utilización de un transpondedor de 40 MHz para el acceso múltiple con señales PSK de  $N$  fases, donde  $\phi$  se refiere al número de fases<sup>18</sup>(Informe 708). La abscisa es la suma en decibels de la EIRP del satélite y la razón G/T de la estación terrestre, que es directamente proporcional a  $E_b/N_0$ .

Si se desea mayor información, en la referencia 18, informe 708, se describen las hipótesis para estas curvas, que indican una ligera ventaja del TDMA en comparación con el FDMA.

En la figura 49-78 se ilustra conceptualmente el TDMA utilizando satélites. Desde distintas estaciones se envían ráfagas de tráfico, en segmentos de tiempo preasignados. A las ráfagas se les asigna un segmento de tiempo específico, y el satélite las retransmite a la Tierra en su orden prescrito. En la figura 49-79 se muestra un típico formato de cuadros de ráfagas TDMA. Entre los factores críticos se incluyen la recuperación de la portadora, recuperación del reloj de

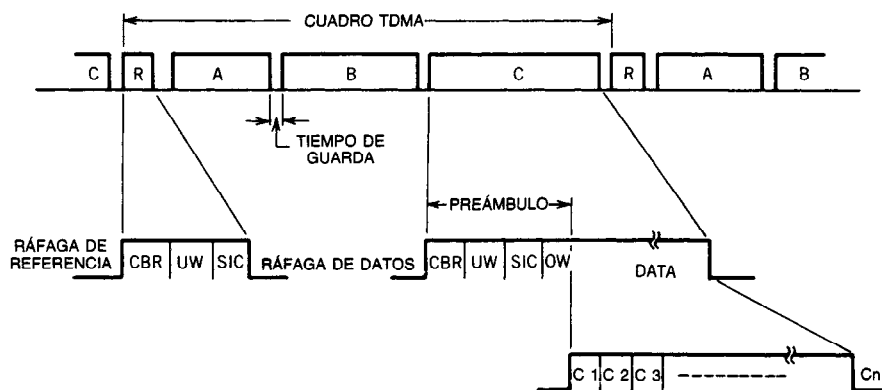


Fig. 49-79. Ejemplos de formato de cuadros y ráfagas de acceso múltiple por división de tiempo (TDMA). CBR = recuperación de la portadora y del reloj; UW = palabra única; SIC = código de identificación de la estación; OW = cable de órdenes; DATA = canal de comunicación; C = canal. Tomado de Miya<sup>10</sup>, con permiso.

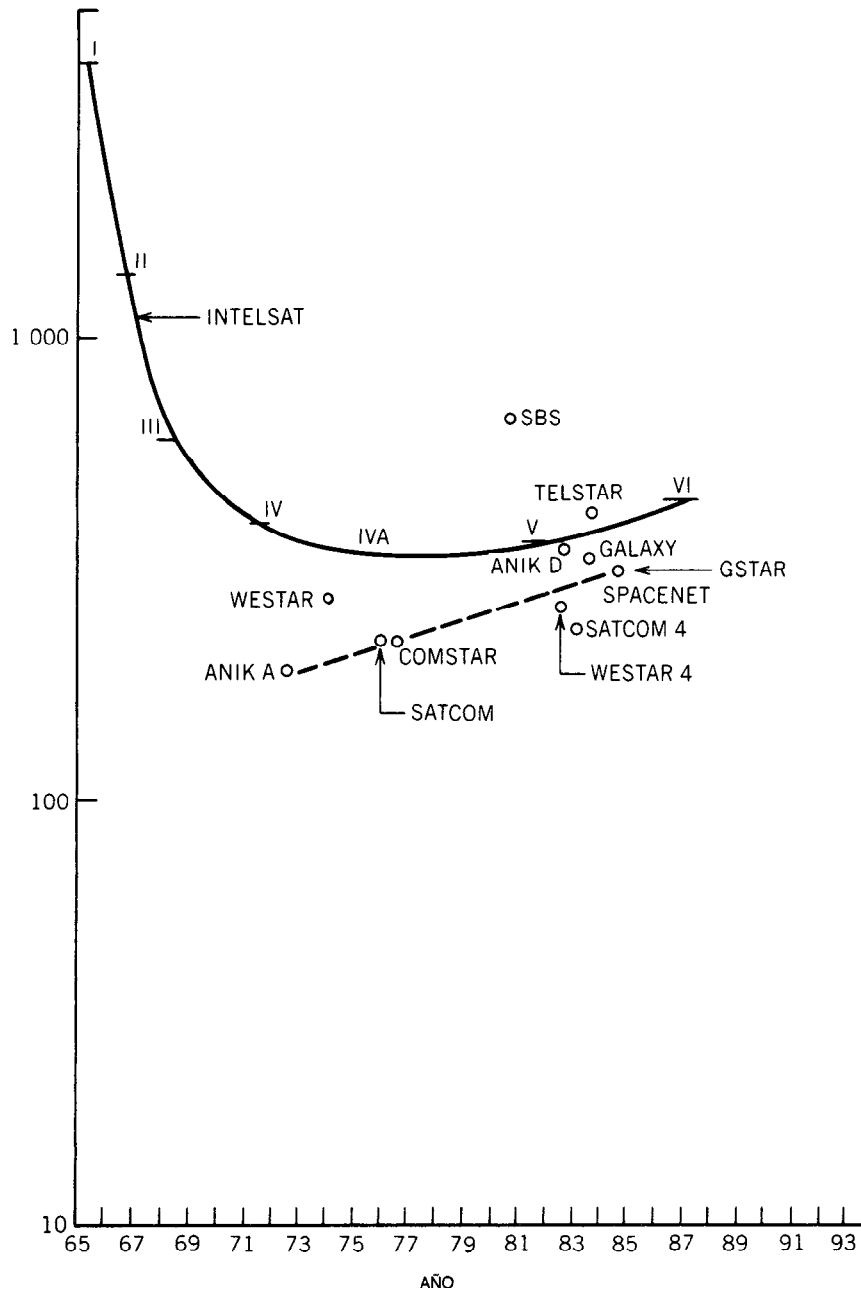


Fig. 49-80. Costo anual de la inversión en el segmento espacial por transpondedor equivalente (ancho de banda de 36 MHz) de los DOMSATS estadounidenses y de los INTELSAT.

sincronización de bits, detección de palabra única, sincronización de ráfagas y adquisición. Esto impone unos requisitos precisos de seguimiento para el sistema. Para múltiples haces puntuales en el satélite, es posible utilizar una matriz de conmutación a bordo a fin de interconectar los haces de los enlaces ascendente y descendente sobre una base de cuadro por cuadro. Esto se efectúa de forma simultánea para todos los

haces, de modo que para  $N$  haces de enlace ascendente y descendente se utiliza una matriz de  $N \times N$ . El conmutador debe ser capaz de funcionar a las velocidades de cuadro TDMA. La referencia 59 contiene una excelente descripción de los sistemas TDMA para satélites, y en la referencia 60 también se describe una variedad de métodos de acceso que no se han examinado en esta sección.

Tabla 49-16. Costos de transpondedores (millones de dólares) de satélites domésticos estadounidenses<sup>7</sup>

Año de adquisición	Satélite	Vehículo espacial +	Lanzamiento =	Vehículo espacial en órbita	Núm. de XPRDR <sup>a</sup> por vehículo espacial	Duración de diseño	Costo anual por XPRDR	Costo anual por XPRDR (\$ de 1982)
1972	WESTAR 1, 2, 3	11	11	22	12	9	0.20	0.50
1973	COMSTAR 1, 2, 3, 4	16	21	37	24	7	0.22	0.52
1974	SATCOM 1, 2, 3	23	15	38	24	8	0.23	0.50
1978	SBS 1, 2, 3, 4	23	24	47	11	7	0.61	0.93
1980	TELSTAR 1, 2, 3	44	30	74	24	10	0.31	0.39
1980	WESTAR 3, 4	25	30	55	24	9	0.25	0.31
1981	GSTAR 1, 2, 3	44	30	74	24	10	0.31	0.34
1981	SATCOM 3R, 4	24	30	54	24	10	0.23	0.26
1981	Galaxy 1, 2, 3	37	30	67	24	8.5	0.33	0.37
1982	Spacenet 1, 2	40	30	70	36	7	0.28	0.28
	Average			54	23	9	0.30	0.44

Fuente: Lovell y Fordyce<sup>8</sup>, reimpresso con permiso.  
<sup>a</sup> XPRDR, transpondedor (transmisor/respondedor).

## 49.4.7 Temas diversos

**Costos de un sistema de comunicación por satélite**

Como ya se observó para los satélites INTELSAT, las mejoras tecnológicas han permitido una continua disminución de costos de servicios de los satélites, desde el primer sistema hasta los actuales. Lovell y Fordyce<sup>8</sup> han dirigido una investigación sobre los sistemas de satélites, incluyendo los costos de lanzamiento, y han definido un factor de costo de mérito, que es el costo de la inversión en órbita por transpondedor equivalente de 36 MHz por año. En la tabla 49-16<sup>8</sup> se presentan los costos del vehículo espacial y su lanzamiento para varios satélites domésticos de Estados Unidos y, para las duraciones dadas de los vehículos, se demuestra que el costo promedio por transpondedor por año (dólares de 1982) es 440 000 dólares. En la figura 49-80<sup>8</sup> se muestra el costo de la inversión en el segmento espacial por transpondedor para los satélites DOMSAT e INTEL-SAT en dólares del año real. Se espera que el factor de mérito para los satélites de radiodifusión directa (DBS, de *direct broadcast satellites*) sea mucho mayor, tal vez hasta de \$2 millones por transpondedor anualmente.

**Utilización de la órbita y política**

La órbita geoestacionaria y el espectro de frecuencias son recursos naturales limitados que no es posible incrementar. La gran proliferación mundial de los sistemas de satélites ha dado una relevancia creciente a las cuestiones técnicas, políticas y legales relacionadas con la utilización eficiente de la órbita<sup>61</sup>. La International Telecommunications Union (ITU), con la asesoría de la CCIR, regula los recursos de órbita y espectro.

Aunque en la figura 49-64 solamente se muestran dos satélites en órbita, dicha figura puede generalizarse a fin de incluir muchos satélites adyacentes y otras localizaciones de estaciones terrestres. El límite inferior en la separación orbital en grados de longitud está dado por la interferencia entre sistemas, lo que a su vez se relaciona con los niveles de potencia, modulación, polarización de antena y niveles de los lóbulos laterales. Además, una fuente extra de interferencia puede ser el compartimiento de bandas de frecuencias con redes terrestres, como es el caso de las bandas de 6/4 GHz. El informe 455-2 de la CCIR aborda las restricciones para el compartimiento, y en las International Radio Regulations se establecen requisitos para que las administraciones se coordinen entre sí con respecto a los nuevos servicios.

La recomendación 466 de la CCIR describe el método del "incremento en la temperatura de ruido" ( $\Delta T$ ) para estimar los efectos de la interferencia. Para el caso en que sistemas con el potencial de interferirse compartan la misma banda de frecuencia, este método implica los pasos siguientes. Primero, es necesario calcular la densidad de flujo  $I_u$  (W/Hz) de interferencia en el receptor del satélite, debida a una estación terrestre interferente. La ganancia de la estación terrestre, su densidad de potencia espectral (W/Hz) y la respuesta de ganancia de la antena del satélite deben incorporarse en las ecuaciones de enlace. Después se convierte la densidad de flujo en un incremento en la temperatura

de ruido del satélite,  $T_s$ , conforme a  $10 \log(k\Delta T_s) = I_u$ . Para el enlace descendente se utiliza un razonamiento parecido a fin de calcular el  $\Delta T_e$  en la estación terrestre debido a la interferencia causada por un satélite. El cambio en la temperatura de ruido para el enlace es  $\Delta T/T = \gamma\Delta T_s + T_e$ , donde  $\gamma$  es la ganancia de transmisión del enlace del satélite, evaluada desde la antena receptora deseada del satélite hasta la antena receptora deseada en la Tierra. Si  $\Delta T/T$  es menor del 4%, entonces la interferencia se considera aceptablemente baja. Si es mayor, es necesario efectuar una coordinación.

**REFERENCIAS BIBLIOGRÁFICAS**

1. A. C. Clarke, "Extra-Terrestrial Relays", *Wireless World* 51(10):303-308.
2. J. R. Pierce, "Orbital Radio Relays", *Jet Propulsion* 25:153-157.
3. J. R. Pierce y R. Kompfner, "Transoceanic Communication by Means of Satellite", *Proc. IRE* 47:372.
4. L. Jaffe, *Communications in Space*, Holt, Rhinehart y Winston, Nueva York.
5. J. R. Pierce, *The Beginnings of Satellite Communications*, San Francisco Press, San Francisco.
6. T. Pirard, "INTERSPUTNIK: The Eastern Brother of INTEL-SAT", *Satellite Commun.* 6(8):38-44.
7. W. Morgan, "Satellite Notebook #31", *Satellite Commun.* 7(13):28-31.
8. R. R. Lovell y S. W. Fordyce, "A Figure of Merit for Competing Communications Satellite Designs", *Space Commun. Broadcast.* 1(1):53-63.
9. Actas Finales de World Administrative Radio Conference, Ginebra.
10. K. Miya, ed., *Satellite Communications Technology*, KDD Engineering and Consulting, Tokyo.
11. R. M. L. Baker, Jr. y M. W. Makenson, *An Introduction to Astrodynamics*, Academic Press, Nueva York.
12. S. Herrick, *Astrodynamics*, Vols. 1 y 2, Van Nostrand Reinhold, Nueva York.
13. H. E. Rowe y A. A. Penzias, "Efficient Spacing of Synchronous Communication Satellites", *Bell Syst. Tech. J.*, pp. 2379-2433.
14. D. V. Z. Wadsworth, "Longitude Reuse Plan Doubles Communication Satellite Capacity of Geostationary Arc", AIAA 80-0507, pp. 198-204.
15. V. J. Slabinski, "Expressions to the Time-Varying Topocentric Direction of a Geostationary Satellite", *COMSAT Tech. Rev.* 5(1):1-14.
16. E. Fthenakis, *Manual of Satellite Communications*, McGraw-Hill, Nueva York.
17. R. G. Gould y Y. F. Lum, *Communications Satellite Systems: An Overview of the Technology*, IEEE Press, Nueva York.
18. *Recommendations and Reports of the CCIR*, Vol. IV, Parte 1, Ginebra.
19. J. J. Spilker, *Digital Communications by Satellite*, Prentice-Hall, Englewood Cliffs, NJ.
20. R. G. Jahn, *Physics of Electric Propulsion*, McGraw-Hill, Nueva York.



21. D. E. Fritz, R. L. Sackheim y H. Macklis, "Trends in Propulsion Systems for Communications Satellites", *Space Commun. Broadcast.* 1(2):173-188.
22. G. W. Durling, "High Power and Pointing Accuracy From Body-Spun Spacecraft", *Space Commun. Broadcast.* 1(1):65-71.
23. *PAM-D / PAM-DII User's Manual*, McDonnell Douglas Astronautics, Huntington Beach, CA.
24. *Ariane 4 User's Manual*, Arianespace, Evry, Francia.
25. *Space Transportation System User's Handbook*, NASA, Washington, DC.
26. P. S. Bargellini y S. J. Campanella, "Overview of Satellite Transmission Techniques", en J. T. Manassah, ed., *Innovations in Telecommunications*, Academic, Nueva York.
27. K. Feher, *Digital Communications-Satellite / Earth Station Engineering*, Prentice-Hall, Englewood Cliffs, NJ.
28. CCIR, *Recommendations and Reports of the CCIR*, Vol. V, Ginebra.
29. J. O. Laws y D. A. Parsons, "The Relation of Raindrop Size to Intensity", *Trans. Am. Geophys. Un.* 24:452-460.
30. D. C. Hogg y T. S. Chu, "The Role of Rain in Satellite Communications", *Proc. IEEE* 63:1308-1331.
31. W. L. Flock, "Propagation Effects on Satellite Systems at Frequencies Below 10 GHz", NASA Ref. Publicación 1108.
32. L. J. Ippolito, R. D. Kaul y R. G. Wallace, "Propagation Effects Handbook for Satellite Systems Design—A Summary of Propagation Impairments on 10- to 100-GHz Satellite Links with Techniques for Systems Design", NASA Ref. Publicación 1082(03).
33. P. Bargellini, ed., "The INTELSAT IV Communications System", *COMSAT Tech. Rev.* 2(2):437-572.
34. R. W. Kreutel y colaboradores, "Antenna Technology for Frequency Reuse Satellite Communications", *Proc. IEEE* 65(3):370-378.
35. D. DiFonzo, A. E. Williams y W. S. Trachtman, "Adaptive Polarization Control for Satellite Frequency Reuse Systems", *COMSAT Tech. Rev.* 6(2).
36. H. L. Van Trees, ed., *Satellite Communications*, IEEE Press, Nueva York.
37. E. I. Podraczky y J. L. Dicks, "International System Development", reimpresión J. T. Manassah, ed., *Innovations in Telecommunications*, Parte B, Academic Press, Nueva York.
38. J. L. Dicks y M. P. Brown, Jr., "INTELSAT IV—A Satellite Transmission Design", *COMSAT Tech. Rev.* 5(1):73-104.
39. R. J. Rusch, J. T. Johnson y W. Baer, "INTELSAT V Design Summary", AIAA 7<sup>a</sup> Communications Satellite Systems Conference, San Diego, CA, pp. 8-20.
40. A. Ghais y colaboradores, "Summary of INTELSAT VI Communications Performance Specifications", *COMSAT Tech. Rev.* 12(2):413-429.
41. R. M. Gagliardi, *Satellite Communications*, Lifefit Learning Publications, Belmont, CA.
42. J. Napoli, "GSTAR—A High Performance Ku-Band Satellite for the 80's", AIAA 9<sup>a</sup> Communications Satellite Systems Conference, pp. 436-447.
43. W. Morgan, "Satellite Notebook #16", *Satellite Communications* 6(7):62.
44. A. Al-Mashat, "The Arab Satellite Communications System", AIAA 9<sup>a</sup> Communications System Conference, pp. 187-191.
45. R. R. Lovell y C. L. Cuccia, "Type-C Communication Satellites Being Developed for the Future", *Microwave Sys. News*, pp. 58-81.
46. I. S. Haas y A. T. Finney, "The DSCS III Satellite: A Defense Communication System for the 80's", AIAA 7<sup>a</sup> Communications Satellite Systems Conference.
47. J. B. Schultz, "Milstar to Close Dangerous C<sup>3</sup>I Gap", *Defense Electr.* 15(3):46-59.
48. T. Takahashi, "The INMARSAT System and Its Development", AIAA 9<sup>a</sup> Communications Satellite Systems Conference, pp. 202-209.
49. D. W. Lipke y colaboradores, "MARISAT—A Maritime Satellite Communications System", *COMSAT Tech. Rev.* 7(2):351-391.
50. D. W. Harris y W. O. Macoughtry, "Tracking and Data Relay Satellite System (TDRSS): A Worldwide View From Space", AIAA N.<sup>o</sup> 1AF-83-78.
51. L. M. Keane, ed., "A Direct Broadcast Satellite System for the United States", *COMSAT Tech. Rev.* 11(2):195-265.
52. G. E. Mueller y E. R. Spangler, *Communication Satellites*, Wiley, Nueva York.
53. A. W. Rudge y colaboradores, eds., *The Handbook of Antenna Design*, P. Peregrinus, Nueva York.
54. R. Strauss, J. Bretting y R. Metirier, "Traveling Wave Tubes for Communications Satellites", *Proc. IEEE* 65:387-400.
55. A. E. Atia y A. E. Williams, "New Types of Waveguide Bandpass Filters for Satellite Transponders", *COMSAT Tech. Rev.* 1(7):21-43.
56. A. E. Atia y A. E. Williams, "Waveguide Filters for Space—Smaller, Lighter—and Fewer Cavities", *Microwave Sys. News* 6(4):75-77.
57. G. Hyde, R. W. Kreutel y L. V. Smith, "The Unattended Earth Terminal Multiple Beam Torus Antenna", *COMSAT Tech. Rev.* 4(2):231-262.
58. *Transmission Systems for Communications*, Bell Telephone Laboratories, Inc.
59. S. J. Campanella y D. Shaefer, Time Division Multiple Access System (TDMA), en K. Feber, ed., *Digital Communications Satellite Earth Station Engineering*, Prentice-Hall, Englewood Cliffs, NJ.
60. J. Martin, *Communications Satellite Systems*, Prentice-Hall, Englewood Cliffs, NJ.
61. D. M. Jansky y M. C. Jeruchim, *Communication Satellites in the Geostationary Orbit*, Artech House, Dedham, MA.

## UNDÉCIMA PARTE

---

# **Sistemas de telemetría, navegación y aterrizaje**



## CAPÍTULO 50

# Principios de navegación

**Emil R. Schiesser**

NASA Johnson Space Center  
Houston, Texas

### INTRODUCCIÓN

**50.1 Definiciones de funciones**

**50.2 Propagación del estado**

**50.3 Determinación del estado**

### **50.4 Observación y medición**

50.4.1 Tipos de medición

50.4.2 Consideraciones de sincronización

### **50.5 Sistemas de navegación**

### **50.6 Resumen**

---

## INTRODUCCIÓN

La **navegación** puede definirse como el proceso de dirigir el movimiento de un vehículo de una posición a otra. En los sistemas de navegación modernos se emplean diversos dispositivos y sistemas electrónicos para realizar las tareas requeridas. Estos sistemas varían considerablemente en complejidad y costo, pero los principios y funciones básicos son comunes.

### 50.1 DEFINICIONES DE FUNCIONES

Entre las funciones que se realizan durante el recorrido en apoyo al proceso de navegación se cuentan:

- **Estimación del estado.** La estimación del estado es la conservación de un conocimiento de la posición de un vehículo y de otros parámetros que describen adecuadamente la situación o estado con fines de navegación. El estado de navegación incluye elementos traslacionales, rotacionales y ambientales esenciales del vehículo y características de éste, tales como la posición, velocidad, aceleración, orientación, frenaje atmosférico y sesgos en la medición del sensor de navegación. La estimación del estado incluye la determinación del estado a través del uso de observaciones a partir de auxiliares de la navegación, y la propagación o extrapolación en el tiempo de esta información en ausencia de dichas observaciones. Esta función se denomina a veces "navegación" en un sentido estrecho.
- **Gobierno.** El gobierno o guiaje es una función por medio de la cual se determina la acción que se requiere para dirigir el vehículo en la trayectoria deseada. Las órdenes de gobierno se basan en una comparación de las condiciones de objetivo proyectadas y deseadas.

- **Control.** El control es la determinación y ejecución de instrucciones de mando detalladas que se dan a dispositivos de forzamiento con base en los resultados de la función de gobierno y el estado de aquellos dispositivos. Estos mandos modifican las fuerzas que actúan sobre el vehículo a través de cambios de orientación de superficies de control o el encendido de chorros de control rotacional y motores cohete traslacionales.
- **Predicción del estado.** Ésta es la predicción del estado futuro con base en estimaciones de estado actuales y el efecto esperado de las fuerzas que actúan sobre el vehículo. La predicción del estado se conoce a menudo como predicción de trayectoria cuando los principales parámetros son la posición y velocidad.
- **Fijación de objetivos.** Ésta es la definición del objetivo del viaje (p. ej., la posición de destino) en términos útiles para el gobierno del vehículo.

### 50.2 PROPAGACIÓN DEL ESTADO

La propagación del estado es la espina dorsal de la estimación del estado y la navegación. Se implementó como "navegación a la estima" con los primeros barcos, y toma la forma de "navegación por inercia" (o inercial) en los sistemas de navegación modernos, tales como los de aviones y misiles.

- **Navegación a la estima.** En la navegación a la estima, la posición del vehículo (barco) se determina con base en la última posición conocida, rumbo o aproamiento, velocidad y tiempo transcurrido desde la última posición (Bowditch).
- **Navegación por inercia.** La navegación por inercia o inercial está asociada con el uso de una unidad o dispositivo de medición inercial para auxiliar la propagación del estado (Pitman). Este

Tabla 50-1. Relaciones entre observaciones y elementos de estado

Tipo de observación	Elementos de estado				
	Traslacionales			Rotacionales	
	Posición	Velocidad	Aceleración	Orientación	Velocidad de cambio
Distancia	x				
Cambio de distancia	x	x			
Cambio de velocidad		x	x		
Ángulo					
Objeto cercano	x				
Objeto lejano				x	
Cambio de ángulo					
Objeto cercano	x	x			
Objeto lejano					x
Presión estática	x				
Presión dinámica	x	x			
Gradiente de gravedad	x				
Ocultación	x	x			

dispositivo opera sobre el principio de que si una masa no experimenta fuerza, permanecerá estacionaria o bien continuará moviéndose con velocidad y dirección constantes hasta que se le aplique una fuerza.

Las unidades de medición inercial (IMU) contienen acelerómetros, que se comportan como péndulos e indican el desplazamiento desde una posición de reposo cuando son perturbados. Se requiere un acelerómetro para cada dirección de movimiento posible. Las IMU también producen mediciones relacionadas con la orientación. Esto se logra con giroscopios, ya sea del tipo electromecánico tradicional o del tipo láser, más reciente. Se puede encontrar más información sobre las IMU en Pitman y en la sección 52.3, referente a sistemas de navegación por inercia de barcos (SINS).

Los acelerómetros no miden fuerzas gravitacionales; por tanto, la propagación del estado mediante las IMU requiere la modelación de tales fuerzas en algunas aplicaciones, tales como los vuelos de misiles. Por ejemplo, un cohete que se sostiene justo sobre el suelo al comenzar a elevarse experimenta una fuerza del cohete que es aproximadamente igual pero opuesta en dirección a la fuerza gravitatoria. Un acelerómetro debidamente alineado en la vertical mediría un cambio de velocidad (aceleración) de alrededor de  $9.8 \text{ m/s}^2$  producido por la fuerza de contacto del cohete, aunque no esté ocurriendo tal cambio de velocidad.

La propagación de la posición y la velocidad se realizan en computadoras instaladas a bordo por integración numérica de mediciones de cambios de velocidad obtenidos de las IMU y aceleraciones calculadas a partir del modelo de fuerzas gravitacionales.

La estimación del estado y la navegación se basan sólo en la propagación del estado cuando no se dispone de mediciones de observación de características externas. Los sistemas inerciales y la propagación del estado en general acumulan errores en el tiempo, ya que el estado inicial, las mediciones de las IMU y los modelos de fuerzas son imperfectos. Para compensar estas tendencias naturales al error, el estado se ajusta en forma periódica mediante el uso de datos de otras fuentes, tales como los auxiliares electrónicos de la navegación que se analizan en los capítulos siguientes.

### 50.3 DETERMINACIÓN DEL ESTADO

Para determinar la posición y velocidad se miden uno o más parámetros que cambian con las variaciones de posición del vehículo. Lo ideal sería medir de manera directa, económica, exacta, continua y confiable los elementos de estado deseados. La mayor parte de los sistemas miden en realidad parámetros diferentes (pero relacionados) de los elementos de estado de interés, y estos parámetros se relacionan sólo con una parte del estado en un momento dado cualquiera. En consecuencia, se requiere un historial de tiempo de mediciones para determinar la posición y velocidad.

La función de corrección del estado consiste en ajustar la posición, velocidad y otras estimaciones de

elementos de estado de manera que una curva resultante del estado en función del tiempo sea consistente con un conjunto de observaciones. La curva se define a través del proceso de propagación del estado. La propagación y corrección del estado pueden visualizarse como la extracción de una señal de un fondo ruidoso mediante el uso de un filtro.

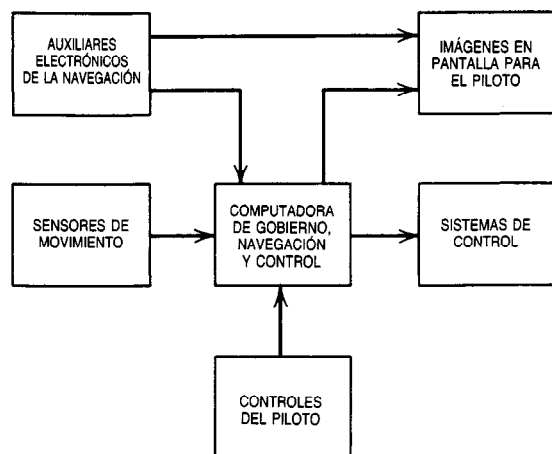
El proceso de ajuste de la curva puede realizarse con un grupo de observaciones, como sucede con un ajuste por mínimos cuadrados de un conjunto de datos fijo. También se aplican técnicas de filtrado secuencial con una observación a la vez o con grupos de observaciones sucesivos. Existen dos técnicas secuenciales básicas, una que implica la inversión de una matriz de las mismas dimensiones que el estado y otra, el filtro de Kalman, que opera sin inversión matricial explícita si las mediciones que se realizan al mismo tiempo no están correlacionadas. De lo contrario se debe invertir una matriz del tamaño de las observaciones correlacionadas localmente. Ambas técnicas tienen que ver con la determinación de la parte del estado observada y de la cantidad para corregir esa parte con base en la incertidumbre presente del estado y la calidad de la observación. La determinación de todo el estado suele requerir el uso de una serie de observaciones en algún intervalo de tiempo, de manera que se disponga de una cantidad suficiente de geometría física y datos para determinar una curva relativamente uniforme a través de las mediciones generalmente ruidosas. Son muy comunes las técnicas de filtrado de Kalman modificadas para el cálculo de la corrección del estado en computadoras instaladas a bordo.

## 50.4 OBSERVACIÓN Y MEDICIÓN

La geometría, tipo y disponibilidad de observación junto con la calidad de la medición, correlación de marcas ("etiquetas") de tiempo y confiabilidad son críticas para la determinación del estado. Las observaciones pueden clasificarse conforme a sus relaciones con los elementos de estado del vehículo, su dependencia respecto a dispositivos y características externos, si se requiere transmisión de energía y si la medición se obtiene a bordo o es externa al vehículo, por ejemplo en tierra. Entre las características externas se cuentan las naturales y artificiales, y los dispositivos de medición pueden ser pasivos, de apoyo, o cooperativos. En la tabla 50-1 se hace un resumen de observaciones y mediciones comunes junto con sus relaciones con los elementos de estado.

### 50.4.1 Tipos de medición

La **distancia** (en algunos casos especiales llamada también "alcance") es la longitud de una trayectoria entre dos puntos, en este caso entre el vehículo y un punto de medición externo. La distancia en línea recta se denomina distancia real (verdadera) u oblicua, y su medición es influida por la refracción de ondas de radio en el medio de propagación. Estos efectos se analizan en la sección 48.2.



**Fig. 50-1.** Diagrama funcional de bloques de un sistema de navegación moderno.

La velocidad de cambio de la distancia se considera instantánea (tiempo infinitesimal). Se determina aproximadamente midiendo la variación de la distancia en intervalos de tiempo breves, a través del desplazamiento Doppler de la radiofrecuencia de la señal de medición de la distancia. La velocidad de cambio de la distancia determinada de esta manera es lo suficientemente exacta en la mayor parte de las aplicaciones de navegación, pero cuando las tareas de navegación resultan complicadas la medición se observa y utiliza como cambio de la distancia. Varios sistemas producen una cuenta de frecuencia Doppler acumulada, de manera

que el usuario pueda seleccionar el intervalo de tiempo para procesar la medición de cambio de la distancia.

Entre las mediciones de ángulos se cuentan el ángulo de llegada de una señal recibida y el ángulo entre dos direcciones rectilíneas ópticas, por ejemplo, una hacia una estrella y la otra dirigida al punto más cercano en el horizonte.

#### 50.4.2 Consideraciones de sincronización

Las observaciones de navegación sólo son válidas en el instante en el cual se miden, y por lo general deben "etiquetarse" en tiempo de forma adecuada para ser útiles. Esto es, se determina un valor del tiempo en el cual se hizo la medición y ese tiempo se asocia con la medición. El proceso de navegación depende de la capacidad de mantener un registro de tiempo adecuado y la sincronización de los datos de medición.

### 50.5 SISTEMAS DE NAVEGACIÓN

En general, un sistema de apoyo del proceso de navegación incluye sensores de movimiento del vehículo, equipo de control del vehículo, equipo de entrada y salida de información, equipo de procesamiento y vinculación (interfaces), y controles manuales. La figura 50-1 es un diagrama funcional de bloques de un sistema de navegación típico que sólo incluye los principales grupos de funciones. Los auxiliares de la navegación electrónicos (que se analizan con detalle en los cuatro capítulos que siguen), junto con los sensores de movimiento de a bordo y entradas del piloto (si el vehícu-

**Tabla 50-2.** Clasificación de los sistemas de navegación

Parámetro de clasificación	Ejemplo	Características y requisitos especiales
Tipo de vehículo	Avión Barco de superficie Submarino Nave espacial	Alta velocidad, larga distancia Baja velocidad, larga duración Baja velocidad, larga duración, comunicaciones difíciles Velocidad y distancia extremas
Área de recorrido	Carriles de tráfico aéreo Áreas de aterrizaje remotas Mar abierto Espacio exterior Zona litoral	Extenso apoyo desde tierra Acceso limitado a sistemas con base en tierra No hay puntos de referencia, se requiere menor exactitud Se requiere exactitud a largo plazo, geometría complicada Se requiere exactitud, se tiene acceso a muchos auxiliares
Base de navegación	A bordo Con base en tierra Mixto o compartido	La estimación del estado se realiza a bordo Requiere enlace de comunicaciones Se transmiten actualizaciones desde tierra firme
Fase de recorrido (ejemplo del avión)	Despegue/ascenso Crucero Aterrizaje	Cambio de estado rápido Cambios menores de la trayectoria planificada Se requiere precisión, cambio de estado rápido

lo es tripulado), proporcionan datos para un subsistema de procesamiento que a menudo se denomina computadora de gobierno, navegación y control (GN&C). En un sistema simple, la computadora GN&C puede ser una calculadora portátil o una regla de cálculo especial. Cualquiera que sea la forma del subsistema de procesamiento, su objetivo consiste en proporcionar información al piloto en forma de imágenes, diagramas, indicadores, etc., y dar instrucciones al sistema de control.

La composición de un sistema de navegación específico depende de la aplicación. Un avión de alta velocidad, por ejemplo, demanda diferentes funciones al sistema de navegación que un barco en lento movimiento por aguas litorales. Existen muchas maneras de clasificar los sistemas de navegación; algunas se ilustran en la tabla 50-2. Estas clasificaciones no son mutuamente excluyentes, ya que un sistema en particular puede incluirse en varias de las clasificaciones del ejemplo. Más aún, puede requerirse que un sistema determinado opere en diferentes situaciones. Por ejemplo, el sistema de navegación de las aeronaves debe operar en las fases de despegue, crucero y aterrizaje.

## 50.6 RESUMEN

La navegación moderna es un proceso que incluye una diversidad de mediciones, cálculos y actividades especializados para satisfacer las necesidades particulares de la aplicación. De especial interés en este libro son los auxiliares de la navegación electrónicos que realizan las mediciones de los parámetros y muchas de las funciones de visualización en pantalla que son parte integral del proceso de navegación. En los cuatro capítulos siguientes se describirán en forma detallada muchos de esos sistemas y dispositivos.

### LIBROS Y REVISTAS

- N. Bowditch, *American Practical Navigator*, Defence Mapping Agency Hydrographic Center, Publicación n.º 9, Vol. 1
- G. Pitman, *Inertial Guidance*, Wiley, Nueva York.
- *Proceedings of the IEEE*. Número especial sobre sistemas globales de navegación. IEEE Press, Nueva York.





## CAPÍTULO 51

# Radar

**Frank R. Castella**

Johns Hopkins University  
Laurel, Maryland

- |  |   |
|--|---|
| <b>51.1 El problema de la navegación</b>           | <b>51.6 Limitaciones del sistema en cuanto a exactitud de la medición</b> |
| <b>51.2 Principios del radar</b>                   | <b>51.7 Efectos de la trayectoria de propagación</b>                      |
| <b>51.3 Tipos de radar</b>                         | 51.7.1 Trayectoria múltiple   |
| <b>51.4 Ecuaciones de funcionamiento del radar</b> | 51.7.2 Refracción   |
| <b>51.5 Resolución del radar ideal</b>             | 51.7.3 Atenuación y dispersión  |
|  | <b>51.8 Radar Doppler y mediciones de velocidad</b>                       |
|  | <b>51.9 Técnicas avanzadas (ondas milimétricas)</b>                       |

### 51.1 EL PROBLEMA DE LA NAVEGACIÓN

La navegación de superficie es el proceso de dirigir el movimiento de un barco de un punto a otro sobre la superficie de la Tierra. El pilotaje, o navegación observada, es la navegación en aguas restringidas e implica la determinación frecuente de la posición relativa a puntos geográficos con un alto grado de exactitud. En la navegación, el radar puede servir no sólo para determinar posiciones, sino también como auxiliar en la evasión de colisiones. Es un auxiliar muy importante para la navegación porque proporciona la misma información durante la noche y con niebla que en condiciones más favorables. Una gran ventaja del radar es que no requiere la cooperación de otras estaciones, a diferencia de otros auxiliares de la navegación.

Para el océano abierto los principales sistemas de navegación son el LORAN, el OMEGA y el sistema de navegación por satélite auxiliar. En el LORAN u OMEGA, las diferencias de tiempo o fase en el receptor del navegante se miden en transmisiones realizadas desde estaciones geográficamente distribuidas. A partir de esta información se puede determinar la posición del navegante. En el sistema de satélite auxiliar de la navegación (navegación por satélite), la posición se determina a partir del historial de frecuencia Doppler de una señal recibida desde un satélite. El cometido del radar en un barco en el océano abierto (es decir, a una larga distancia de tierra) consiste principalmente en evitar colisiones con otros barcos, icebergs, islas, etc. En la proximidad de tierra el radar puede funcionar como un sensor de precisión para la localización, además de su función normal de evasión de colisiones. Con mapas disponibles, la imagen que aparece en la pantalla del radar (el indicador de la posición en el plano o PPI) se puede comparar con el mapa para determinar la ubicación de la nave. Sin embargo, la imagen del radar en muchos casos no tendrá la calidad del mapa debido a factores tales como la resolución del sensor, efectos de ocultación del terreno y propiedades de dispersión de los objetivos, de manera que suele necesitarse el juicio del ser humano para identificar líneas costeras, crestas de montañas, boyas y otras referencias. Es concebible que en equipos futuros se pueda realizar un procedimiento de comparación de mapas, en el que se correlacionarán mapas almacenados internamente con la información PPI del radar para determinar la ubicación del barco. En un sistema de este tipo quizá un operador sólo necesite indicar la ubicación aproximada, y después una computadora realizará la función sucesiva de búsqueda y correlación de mapas. El producto final serían las coordenadas del barco (es decir, latitud y longitud) y su ubicación en la retícula de coordenadas del mapa.

La evasión de colisiones es un área en la cual es obligatorio tener un radar. En medios rodeados por niebla es el único método óptimo para realizar esta función. El operador en la actualidad determina la presencia de otros barcos u obstáculos, y a partir del historial PPI determina si representan una amenaza de colisión. Si se predice una colisión en algún tiempo futuro, se efectúa una maniobra que coloca al barco en

un curso libre de choque. Esta situación se monitoriza continuamente hasta que la separación física y las velocidades relativas indican que ya ha pasado el peligro. El proceso de seguimiento automático de objetivos y el cálculo para la evasión de colisiones se están implementando con dispositivos llamados auxiliares para la evasión de colisiones. Éstos ofrecerán información al capitán o navegante del barco, y éstos podrán utilizar esta información según lo deseen.

### 51.2 PRINCIPIOS DEL RADAR

Un radar opera de la manera siguiente: un transmisor genera una serie de pulsos sinusoidales muy breves que se radian a través de una antena con ancho de banda reducido. Estos pulsos se propagan por el espacio aproximadamente a la velocidad de la luz e interactúan con objetos (blancos u objetivos) como boyas, barcos y aviones. Parte de la energía incidente en estos objetos se refleja en la dirección del radar, es detectada por el receptor del radar y se visualiza en la pantalla de un tubo de rayos catódicos (CRT). Algunos parámetros típicos de un sistema de radar que se emplean con fines de navegación y se ilustran en la figura 51-1 son:

Duración del pulso	$0.1 \times 10^{-6}$ s o $0.1 \mu\text{s}$
Intervalo entre pulsos	$10^{-3}$ s o 1.0 ms
Radiofrecuencia de los pulsos radiados	$10^{10}$ Hz o 10 GHz
Velocidad de rotación mecánica de la antena	15 rpm

El intervalo de tiempo entre la transmisión de un pulso y la recepción de la energía devuelta o eco desde el objeto indica la distancia entre la antena y el objeto. Si la distancia al objeto es  $R$ , el intervalo de tiempo  $T$  está dado por la fórmula:

$$T = \frac{2R}{C}$$

donde  $C$  es la velocidad de la luz. La distancia máxima para la detección inequívoca de un objeto (es decir, la energía del eco regresa antes de la transmisión del siguiente pulso) es:

$$R_{\text{máx}} = \frac{C}{2} T_{\text{máx}}$$

Como  $T_{\text{máx}}$  corresponde al intervalo entre los pulsos, entonces:

$$R_{\text{máx}} = \frac{3 \times 10^8}{2} \text{ m/s} \times 10^{-3} \text{ s} = \frac{3}{2} \times 10^5 \text{ m} = 150 \text{ km}$$

Un objeto a 160 km que refleja una cantidad de energía de eco lo suficientemente grande hacia el receptor del radar se visualizará en el CRT a una distancia de 10 km. Esto se denomina detección ambigua de la distancia.

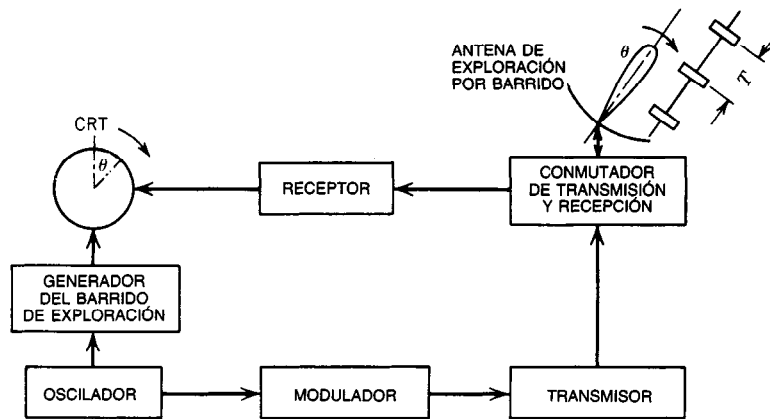


Fig. 51-1. Diagrama de bloques de un sistema de radar. CRT = tubo de rayos catódicos.

La visualización en la pantalla del CRT, también llamada PPI, opera de manera que el eco del objeto se visualiza a lo largo de un radio vector desde el centro del tubo. Para un objeto situado a 75 km, con deflexión de escala completa correspondiente a 150 km, el rebote (eco) del objeto se verá a medio camino a lo largo del radio vector. Por tanto, si el diámetro del tubo es de 35 cm, el eco se presentará a 8.75 cm del centro del tubo.

El ángulo de acimut del radio vector está relacionado electromecánicamente con la dirección de la antena. Si la presentación PPI es tal que la dirección norte corresponde a la parte superior del tubo, entonces un objeto detectado  $45^\circ$  al este del Norte se visualiza en un ángulo  $\theta = 45^\circ$ . El radio vector gira en torno al centro del tubo a la velocidad de rotación mecánica de la antena.

Dado que el radar transmite pulsos de alta potencia por breves periodos de tiempo (bajo régimen de trabajo), la potencia promedio transmitida puede ser bastante baja. Con los parámetros ya definidos, y una potencia máxima (pico) de pulsos de  $10^6$  W (1 MW), la potencia promedio transmitida es  $10^6 \times 10^{-7}/10^{-3} = 100$  W. Los ecos visibles en la pantalla pueden provenir de objetos deseados y no deseados. Es función del observador humano estudiar el CRT para discriminar entre los ecos deseados y los indeseados. Este proceso de detección se puede automatizar si sólo interesan objetivos puntiformes, como barcos, aviones y boyas. Si se desea mapear entradas de puertos y ecos parásitos de tierra, el operador humano es el mejor detector.

### 51.3 TIPOS DE RADAR

Los radares pueden clasificarse por sus características de cobertura y por el tipo de onda de forma que transmiten. También se pueden clasificar como monostáticos o biestáticos, de modo que el primero se refiere a las antenas de transmisión y recepción en el mismo sitio geográfico (y por lo general la misma antena) y el último a las antenas de transmisión y recepción en lu-

gares separados. Los radares pueden ser del tipo de vigilancia (o exploración) tridimensionales (3D) o bidimensionales (2D), donde se abarcan los  $360^\circ$  en acimut sobre una base periódica.

**Radares tridimensionales.** Tienen un ancho de banda estrecho en las coordenadas de acimut y elevación, de modo que con la coordenada de la distancia se pueden determinar las tres coordenadas de un objetivo. La cobertura de elevación suele ser de cerca de  $45^\circ$  (fig. 51-2).

**Radares bidimensionales.** Tienen un haz en abanico estrecho en acimut, y por tanto no pueden determinar el ángulo de elevación del objetivo (fig. 51-3).

El periodo de tiempo típico para abarcar  $360^\circ$  en acimut,  $45^\circ$  en elevación y distancias hasta 200 nm es 4 s. La cobertura en acimut se puede obtener de ma-

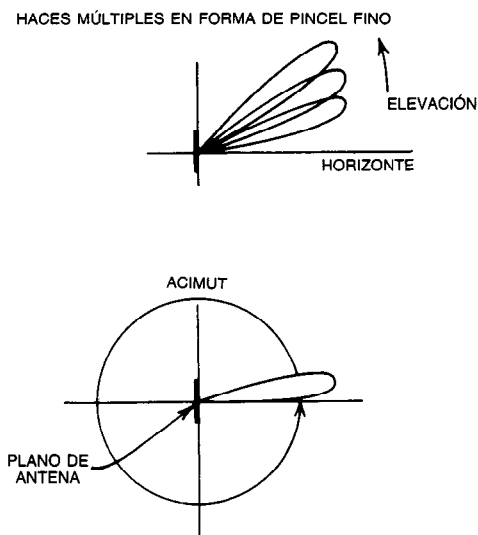


Fig. 51-2. Cobertura del radar tridimensional.

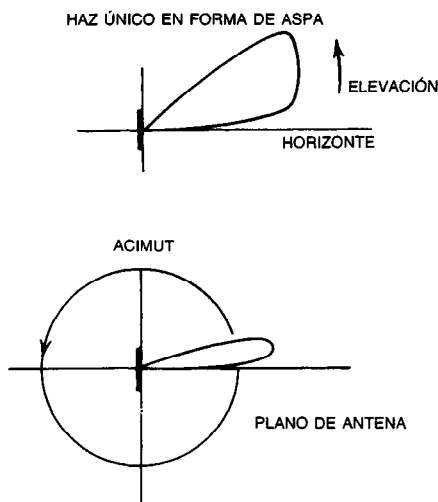


Fig. 51-3. Cobertura del radar bidimensional.

nera más económica con una antena de exploración mecánica, mientras que un arreglo en fase más costoso pero más flexible con caras múltiples puede efectuar el barrido (exploración) en forma electrónica. Un arreglo en fase de tres caras guiado electrónicamente puede abarcar  $360^\circ$  en acimut si se pide que cada cara guíe su haz  $\pm 60^\circ$  en acimut. La cobertura en elevación que logra un radar de vigilancia 3D suele lograrse mediante el barrido de un haz de elevación estrecho cuando la antena gira o se guía en acimut. Este haz de elevación estrecho puede hacerse barrer (explorar) aplicando un gradiente de fase a través de la abertura de la antena en la dirección vertical. La magnitud del gradiente determina el ángulo de elevación al cual se dirige el haz. La función de un radar de vigilancia 2D o 3D consiste en detectar objetivos en un campo de visión extenso. El seguimiento de los objetivos deseados se puede realizar por medio de un sistema automático de seguimiento simultáneo con la exploración, con lo cual las detecciones del radar de vigilancia sirven como entradas de una computadora y los objetivos se siguen vía algún algoritmo matemático.

**Radar de seguimiento.** Es un radar 3D dedicado a un solo objetivo con anchos de haz reducidos en acimut y elevación. Mientras que un radar de vigilancia sólo puede volver a pasar por un objetivo después de algunos segundos, un radar de seguimiento (o rastreo) está dirigido continuamente a un objetivo deseado y por tanto puede determinar con mayor exactitud sus coordenadas. Los anchos de haz comunes de un radar de seguimiento son  $1^\circ$  en acimut y  $1^\circ$  en elevación (fig. 51-4). Para adquirir el objetivo deseado con un haz estrecho como éste se requiere una designación previa desde un radar de vigilancia u otra fuente y la búsqueda de un volumen especificado en el espacio alrededor de ese punto de designación. Por lo general, el rastreo del objetivo deseado en ángulo se realiza aplicando técnicas de rastreo de ángulos que se conocen como barrido cónico o monopolso.

Los radares se pueden clasificar asimismo por el tipo de forma de onda que transmiten. Un radar de onda de forma continua (CW) o modulado en frecuencia (FM) transmite el 100% del tiempo y por lo general necesita antenas de transmisión y recepción independientes para lograr una máxima sensibilidad de detección. Estos radares funcionan con base en el efecto Doppler (sec. 51.8), que se obtiene con objetivos en movimiento. Con la transmisión CW no se puede determinar la distancia al objetivo, pero sí con la transmisión FM. Estos radares pueden detectar objetivos en zonas con muchos ecos parásitos o emborronamiento. Los radares indicadores de objetivos en movimiento (MTI) y Doppler de pulsos también pueden detectar objetivos en movimiento entre ecos parásitos o emborronamiento. Por lo general, en estos tipos de radar se emplean transmisiones pulsátiles y los ecos parásitos se rechazan en forma un tanto menos efectiva que en los radares CW o FM, pero la distancia se determina con mayor exactitud.

En el tipo de radar más común se emplean transmisiones pulsátiles y no se aprovecha el efecto Doppler para el caso de objetivos en movimiento. Estos radares se denominan incoherentes y sólo utilizan las características de amplitud del eco del objetivo.

Los radares en los que se aprovecha el efecto Doppler emplean las características de fase y amplitud del eco del objetivo.

#### 51.4 ECUACIONES DE FUNCIONAMIENTO DEL RADAR

Revisten interés las siguientes características de funcionamiento de un sistema de radar:

1. Distancia de detección de un objeto con área de eco especificada o para una sección transversal del radar dada.
2. Capacidad de resolución en la distancia y ángulo de objetos poco espaciados.
3. Exactitud de localización de objetos en distancia y ángulo.
4. Capacidad de detectar objetos en movimiento en medio de ecos parásitos de mar y tierra con densidad especificada.
5. Entorno de cobertura del radar (es decir, acimut de distancia, límites de elevación).

La distancia de detección de un objeto se puede determinar calculando la razón de señal sobre ruido (SNR) en el receptor de señales reflejadas por el objetivo. Ésta se determina como sigue:

$$SNR = \left( \frac{P_t G_t}{4\pi R^2} \right) \cdot \left( \frac{\sigma}{4\pi R^2} \right) \cdot \left( \frac{G_r \lambda^2}{4\pi} \right) \cdot \left( \frac{1}{kTB} \right)$$

La primera cantidad del lado derecho de la ecuación es la densidad de potencia en watts por metro cuadrado a una distancia de  $R$  metros desde un transmisor de  $P_t$  watts y una antena de ganancia  $G_t$ . El segundo factor es la fracción de la potencia que incide en un objetivo de

sección transversal  $\sigma$  (área aparente), que se rerradia a una distancia de  $R$  metros. El producto de los dos primeros factores es la densidad de potencia que se refleja hacia el punto de transmisión desde el objetivo.

Por otra parte, la tercera cantidad del segundo miembro de la ecuación es el área efectiva de la antena de recepción, que se expresa en términos de la ganancia  $G_R$  y la longitud de onda  $\lambda$ , en metros. El área de la antena se expresa en metros cuadrados. El producto de los tres primeros términos es la potencia de señal total recibida en el receptor, dado que se transmitieron  $P_t$  watts.

Esta sección transversal del radar depende de la composición, forma y tamaño del objeto, longitud de onda  $\lambda$  y polarización de las transmisiones del radar. Estas transmisiones pueden ser polarizadas en forma vertical, horizontal o circular. Un valor de referencia de la sección transversal del radar es  $\sigma_0 = 1 \text{ m}^2$ .

La última cantidad del denominador es el ruido del receptor, que compite con la señal en el proceso de detección.  $k$  es la constante de Boltzmann, igual a  $1.37 \times 10^{-23} \text{ J/K}$ , y  $T$  es la temperatura de ruido en kelvins.  $B$  es el ancho de banda del receptor en hertz y es aproximadamente igual al recíproco de la duración del pulso transmitido.  $F$  es el factor de ruido del receptor (mayor que uno), una medida del ruido introducido en forma inherente por los amplificadores y mezcladores del receptor.

La razón de señal sobre ruido debe ser lo suficientemente grande para detectar un objeto de manera confiable. Este valor requerido es función de las siguientes cantidades:

1. Número de pulsos  $N$  que regresan desde un objetivo cuando la antena rastrea por él.
2. Probabilidad de detección deseada.
3. Probabilidad de falsa alarma deseada.

El punto 1 influye en el proceso de detección, ya que múltiples pulsos se integran a valores de intensidad más grandes en la pantalla del CRT debido a la capacidad de integración del fósforo que envuelve la superficie interior del CRT. Para un ancho de haz en acimut de  $1^\circ$ , velocidad de rotación de la antena de  $90^\circ/\text{s}$  y frecuencia de repetición de pulsos de 1 000 pps, el número de pulsos reflejados por un objeto puntual durante su tiempo de "iluminación" es  $N = 1\,000 \text{ pps} \times$

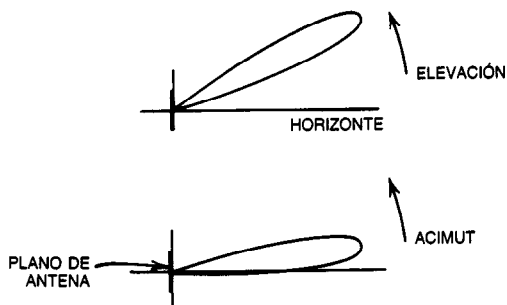


Fig. 51-4. Cobertura del radar de rastreo (seguimiento).

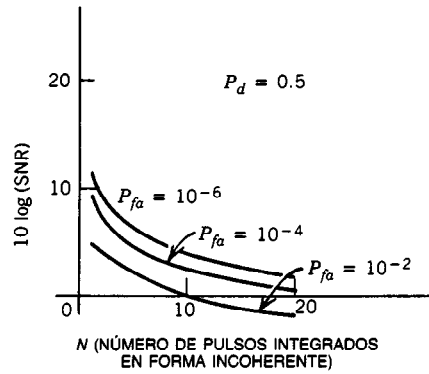


Fig. 51-5. Razón de señal sobre ruido (SNR).  $P_d$  = probabilidad de detección;  $P_{fa}$  = probabilidad de falsa alarma.

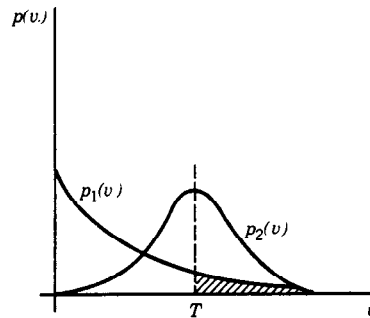


Fig. 51-6. Densidad de probabilidad de voltaje del detector.  $p_1(v)$  = no hay objetivo;  $p_2(v)$  = objetivo presente.

$1^\circ/90^\circ \text{ s} \approx 11$  pulsos. Como el ruido interferente tiene propiedades aleatorias, el proceso de detección de objetivos es un evento (suceso) estadístico. La figura 51-5 es una gráfica de la razón de señal sobre ruido en decibels (es decir,  $10 \log \text{SNR}$ ) requerida para la detección en función del número de pulsos  $N$ . La probabilidad de detección,  $P_d$ , es igual al 50% y las probabilidades de falsa alarma,  $P_{fa}$ , son  $10^{-6}$ ,  $10^{-4}$  y  $10^{-2}$ . Un valor de  $P_{fa} = 10^{-4}$  indica que el ruido será falsamente declarado como el objetivo una vez en 10 000 oportunidades.

Puede apreciarse que la razón de señal sobre ruido requerida aumenta con  $P_{fa}$  decreciente y decrece cuando se integran más pulsos.

En la figura 51-6 se ilustra la naturaleza estadística de la decisión de detección. Las curvas 1 y 2 representan las funciones densidad de probabilidad de voltaje del receptor ( $v$ ) después del detector. El efecto de integración en el fósforo del CRT se ha modelado como la sumatoria de las amplitudes de una sucesión de  $N$  pulsos, como lo haría un detector automático. La función  $p_1(v)$  es la densidad de probabilidad de  $v$  cuando sólo está presente ruido en la entrada.  $p_2(v)$  es la densidad de probabilidad de  $v$  cuando la entrada está presente la señal más el ruido. Puede observarse que  $p_2(v)$  tiene un valor medio mayor que  $p_1(v)$ .

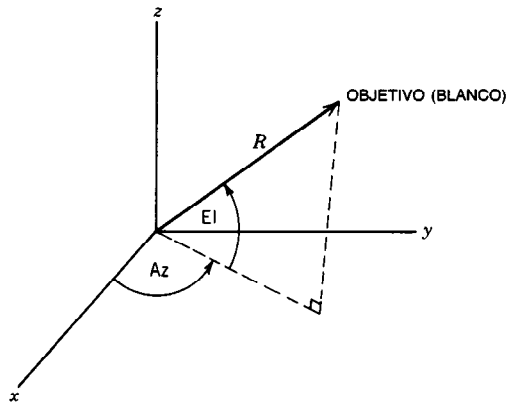


Fig. 51-7. Coordenadas del radar. Az = ángulo de acimut; El = ángulo de elevación; R = distancia.

Si se establece un valor umbral de  $T$  volts como el umbral de detección, los valores  $P_d$  y  $P_{fa}$  están dados por:

$$P_d = \int_T^{\infty} p_2(v) dv$$

$$P_{fa} = \int_T^{\infty} p_1(v) dv$$

De esta forma se pueden obtener resultados como los de la figura 51-5.

### 51.5 RESOLUCIÓN DEL RADAR IDEAL

La **resolución** es la capacidad del radar de distinguir dos objetivos muy próximos entre sí. Los objetivos se pueden resolver en cualesquiera dos de las cuatro coordenadas siguientes: distancia real (verdadera u oblicua), ángulo de acimut, ángulo de elevación y velocidad de cambio de la distancia. Las tres primeras de estas coordenadas de radar se ilustran en la figura 51-7, donde el radar está situado en el origen del sistema de coordenadas  $x, y, z$ . La resolución de la velocidad de cambio de la distancia sólo está disponible para los radares en los que se emplea procesamiento Doppler, mientras que la resolución del ángulo de elevación sólo lo está para radares 3D. En un procesador incoherente ordinario con una antena de haz en abanico (2D), la resolución de objetivos muy juntos se puede lograr exclusivamente con las coordenadas de distancia y ángulos de acimut (es decir, el ángulo de rumbo o marcación). La resolución de la velocidad de cambio de la distancia se puede lograr con el procesador Doppler que se describe en la sección 51.8 si los dos objetivos dan respuestas que están en filtros adyacentes. Este requisito se cumple cuando:

$$\Delta \dot{r} > \frac{291.58}{f_0 N T} \quad \text{nudos}$$

donde  $\Delta \dot{r}$  = diferencia de velocidad de cambio de la distancia de los dos objetivos, nudos.

$f_0$  = frecuencia del transmisor, GHz

$N$  = número de pulsos que se utilizan en la transformada rápida de Fourier

$T$  = separación entre pulsos adyacentes, ms

Por tanto, si se transmite un tren de pulsos de 1 kHz a  $f_0 = 10$  GHz y se emplea una transformada rápida de Fourier de ocho puntos, entonces dos objetivos con diferencias de velocidad de cambio de la distancia de 36 nudos producirán respuestas de filtro separadas por un ancho de pulso. Por tanto, la resolución de la velocidad de cambio de la distancia es de 36 nudos en este caso.

La resolución de la distancia se logra cuando los dos objetivos están separados en distancia dos anchos de pulso. Como un ancho de pulso  $\tau$  da una respuesta en un intervalo de distancia  $\Delta r$  para un objetivo puntual, donde:

$$\Delta r = \frac{C}{2} \tau$$

la resolución de la distancia está dada por:

$$\text{Resolución de la distancia} \approx 2\Delta r = C\tau$$

Con un ancho de pulso transmitido de  $0.1 \mu s$ , la resolución de la distancia es  $3 \times 10^8 \text{ m/s} \times 0.1 \times 10^{-6} \text{ s} = 30 \text{ m}$ . Por tanto, objetivos separados 30 m en distancia se pueden resolver con transmisiones de  $0.1 \mu s$ . Asimismo, se pueden obtener anchos de pulso efectivos estrechos codificando (modulando) pulsos de más larga duración. Por ejemplo, si se transmiten pulsos de duración  $\tau$  y el ancho de banda modulado de cada pulso es  $\Delta f$ , entonces la duración de pulso efectiva después de la decodificación en el receptor es:

$$\tau_{ef} = \frac{\tau}{\tau \Delta f} = \frac{1}{\Delta f}$$

De esta forma, si cada pulso tiene un ancho de banda de modulación igual a 10 MHz, entonces el ancho del pulso efectivo después de la decodificación en el receptor es  $0.1 \mu s$ . Esta codificación se puede realizar con pulsos barridos linealmente en frecuencia (es decir, compresión de pulsos) y modulaciones codificadas en fase o frecuencia. Con las limitaciones de potencia pico del transmisor, de esta forma se puede obtener más potencia promedio. La desventaja de codificar los pulsos individuales es que tienen que decodificarse en el receptor, lo cual lo hace más complicado. Asimismo, las transmisiones de duración más larga incrementan la distancia mínima a la cual se pueden detectar objetivos, ya que el receptor está apagado cuando el transmisor está encendido.

La resolución angular de un sistema de radar está limitada por el ancho del haz del patrón de la antena. Este ancho del haz está relacionado con el tamaño físico de las longitudes de onda de la abertura de la antena (es decir, el área radiante), como se ilustra en la figura 51-8. Si las dimensiones de la abertura son  $L_2$

metros de ancho por  $L_1$  metros de altura, entonces el ancho del haz de 3 dB en acimut (es decir, entre los puntos de potencia mitad del patrón de transmisión) es aproximadamente  $\lambda/L_2$ , mientras que el ancho del haz de elevación de 3 dB es aproximadamente  $\lambda/L_1$  radianes. Por tanto, una abertura de  $10\lambda$  produce un ancho del haz de  $1/10$  radián o  $5.6^\circ$ . Suele suponerse una separación de dos anchos del haz como el límite de resolución del ángulo.

Los límites de resolución teóricos son un tanto menores que los valores que se dan aquí, que son límites de "valores de diseño". Por ejemplo, el límite de Rayleigh de la resolución angular es:

$$\Delta\theta = 1.22 \frac{\lambda}{D}$$

donde  $\lambda$  es la longitud de onda y  $D$  es la dimensión lineal de la abertura. Sin embargo, esto se aplica a objetivos de igual intensidad y objetivos de dispersión incoherente. En el caso de dispersores desiguales, la respuesta de lóbulo lateral de un dispersor puede encubrir la respuesta del lóbulo principal del objetivo más débil. Existe una situación similar en la resolución de la distancia, en especial cuando se utilizan pulsos codificados. Al decodificar estos pulsos se generan lóbulos laterales de tiempo, que interfieren en el proceso de resolución. Se pueden obtener límites de resolución de la velocidad de cambio de la distancia menores que el valor que se presentó antes si se aprovecha el conocimiento de las funciones de transferencia de los filtros. La información de amplitud de las respuestas en los filtros adyacentes puede utilizarse para determinar dónde está el objetivo en el intervalo Doppler. Cuando se intentan afinar estos límites de resolución, la presencia de ruido genera un límite inferior impenetrable.

### 51.6 LIMITACIONES DEL SISTEMA EN CUANTO A EXACTITUD DE LA MEDICIÓN

En la sección 51.5 se definió la resolución como la capacidad de distinguir dos objetivos muy próximos entre sí. En esta sección se analizan las exactitudes que se pueden lograr con un sistema de radar cuando se

miden las coordenadas de un solo objetivo: distancia real, ángulo de acimut, ángulo de elevación y velocidad de cambio de la distancia. Las exactitudes suelen ser una fracción de los valores de resolución y son función del método de medición, la onda de forma transmitida y la cantidad de ruido presente. También se analizan en esta sección las exactitudes que se pueden lograr cuando se realiza el seguimiento. Este término se refiere a la combinación de mediciones sucesivas de un objetivo a fin de estimar en forma más exacta su trayectoria. De esta forma se puede obtener un rendimiento mejor, ya que los efectos del ruido se reducen cuando se combinan múltiples mediciones.

La exactitud de una sola medición es inversamente proporcional a la raíz cuadrada de la razón de señal sobre potencia de ruido. Se darán ejemplos de esto para el caso de la exactitud de medición de la frecuencia Doppler y la exactitud de medición de la posición angular. Las componentes en fase y cuadratura ( $I$  y  $Q$ ) de una señal combinada con ruido son:

$$I = a \cos \theta + \eta_I$$

$$Q = a \sin \theta + \eta_Q$$

donde  $a$  = amplitud de la señal

$\theta$  = fase de la señal

$\eta_I, \eta_Q$  = dos componentes del ruido

Cuando el objetivo tiene frecuencia Doppler  $f_d$ , la fase de la señal está dada por:

$$\theta = \Pi f_d t + \theta_0$$

donde  $\theta_0$  es una fase inicial arbitraria. Un circuito discriminador o de medición de la frecuencia, estima la frecuencia Doppler  $f_d$  como:

$$\hat{f}_d = \frac{1}{2\Pi T} \left[ \tan^{-1} \left( \frac{Q_2}{I_2} \right) - \tan^{-1} \left( \frac{Q_1}{I_1} \right) \right]$$

donde  $T$  = duración de la señal, durante la cual la frecuencia Doppler es constante

$I_1, Q_1$  = componentes en cuadratura de la entrada en el tiempo  $t_1$

$I_2, Q_2$  = componentes en cuadratura en el tiempo  $t_2 = t_1 + T$

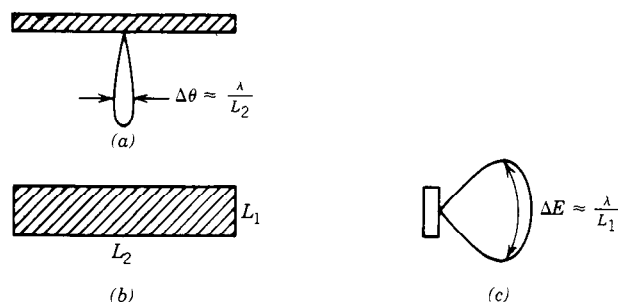


Fig. 51-8. Anchos de haces de antenas para una abertura rectangular: a) vista superior, b) vista frontal, c) vista lateral.



Cuando se determina la desviación estándar de esta expresión para ver cuán exacta es la estimación, se obtiene:

$$\sigma_{f_d} = \frac{1}{2\pi T \sqrt{a^2/2\sigma_n^2}} = \frac{1}{2\pi T \sqrt{\text{SNR}}} = \frac{K}{T \sqrt{\text{SNR}}} \text{ Hz}$$

donde  $a^2/2\sigma_n^2$  es la razón de entrada de señal sobre ruido, SNR. Así, la exactitud de medición de la frecuencia Doppler es inversamente proporcional a la duración de la señal  $T$  e inversamente proporcional a la raíz cuadrada de la razón de entrada de señal sobre ruido. Todas las técnicas de medición de la frecuencia tienen la misma relación funcional con SNR, donde la única diferencia es el valor de  $K$ , que en este ejemplo es  $1/2 \pi$ .

Se puede obtener un resultado similar con un método monopulso de comparación de fases para medir la dirección de llegada de un frente de ondas de señales. Monopulso significa que esta medición se puede realizar con un solo pulso. En este método, que se ilustra en la figura 51-9, dos antenas separadas una distancia  $D$  y con el mismo patrón de respuesta angular  $f(\theta)$  reciben las señales siguientes:

$$S_1 = f(\theta)\cos(\omega_0 t + \psi) + n_1$$

$$S_2 = f(\theta)\cos(\omega_0 t - \psi) + n_2$$

donde  $\psi = (2\pi/\lambda) D \sin \theta$

$n_1, n_2$  = componentes del ruido superpuestas en cada señal

$\lambda$  = longitud de onda de radiación

$\theta$  = dirección de la señal fuera de alineamiento óptico

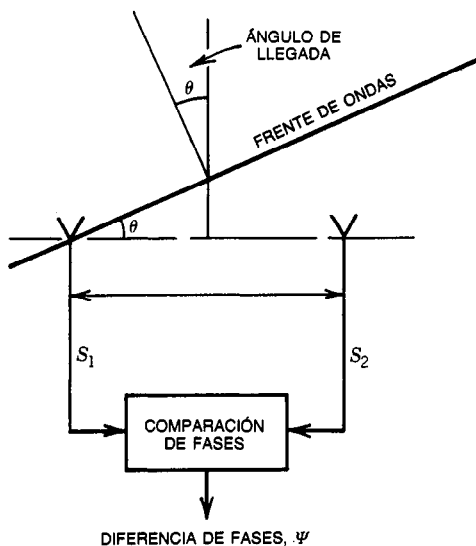


Fig. 51-9. Medición de fase monopulso.

La dirección de alineamiento óptico es perpendicular a la línea que conecta las dos antenas. A partir de la medición del ángulo de fase,  $\psi$ , se puede estimar el ángulo de llegada (arriba) de la señal:

$$\hat{\theta} = \sin^{-1} \frac{\lambda \psi}{2\pi D}$$

El error medio en esta medición es cero, y la desviación estándar es:

$$\sigma_{\theta} = \frac{\lambda}{2\pi D \sqrt{2 \cdot \text{SNR}}}$$

Por tanto, la exactitud de la determinación angular es inversamente proporcional a la raíz cuadrada de la razón de entrada de señal sobre ruido. Por tanto, dos mediciones disimilares, tales como la frecuencia Doppler y el ángulo de llegada, conducen a la misma dependencia respecto a la razón de entrada de señal sobre ruido. Se puede deducir una dependencia semejante cuando se estima la distancia del objetivo.

El resultado anterior para la exactitud del ángulo de llegada supone una medición de tipo interferómetro. En un sistema de radar de exploración en el cual el patrón de la antena barre el objetivo y la posición angular de éste se estima por medio del centroide de las respuestas a pulsos individuales, los resultados no son tan exactos. La estimación del ángulo por esta técnica se considera altamente exacta cuando  $\sigma_{\theta}$  es el 10% de  $\lambda/D$ . Se puede decir lo mismo para las mediciones de distancias. Cuando la medición de la distancia  $\hat{r}$  tiene una desviación estándar igual al 10% de la duración del pulso [es decir,  $\sigma_r = 0.1 (C\tau/2)$ ], la medición se considera altamente exacta.

Cuando se combinan mediciones sucesivas en un intervalo de tiempo para estimar en forma más exacta las coordenadas del objetivo, debe considerarse el movimiento del objetivo durante ese intervalo. Si el desplazamiento del objetivo es despreciable durante el tiempo en el cual se colectan  $N$  mediciones, la exactitud resultante  $\sigma_N$  está dada por:

$$\sigma_N = \frac{\sigma_1}{\sqrt{N}}$$

donde  $\sigma_1$  es la desviación estándar de una sola medición.

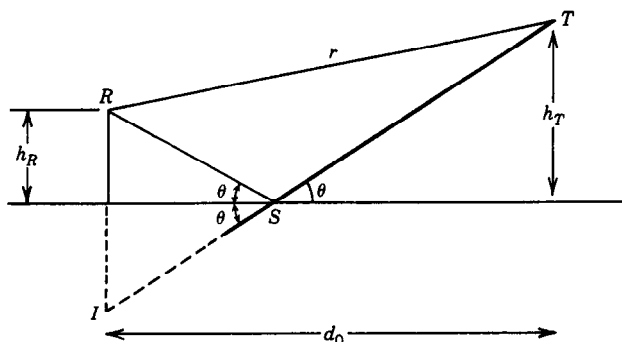
Cuando la velocidad del objetivo es importante durante el periodo de medición y se necesita estimar esta velocidad y también la distancia, la desviación estándar del error en la distancia es:

$$\sigma_N = \frac{2}{\sqrt{N}} \sigma_1$$

En este caso la estimación  $\hat{R}$  es:

$$\hat{R} = \sum \alpha_i [R_0 + V_0(t_i - t_0) + n_i]$$

donde las  $\alpha_i$  se eligen de modo que se minimice el error en un sentido de mínimos cuadrados. Extrapolando



**Fig. 51-10.** Geometría de trayectoria múltiple.  $R$  = ubicación del radar;  $T$  = ubicación del objetivo;  $I$  = ubicación de la imagen del radar;  $S$  = punto de reflexión;  $h_R$  = altura de la antena del radar;  $h_T$  = altura del objetivo;  $r$  = distancia real.

estos resultados al caso en que se tienen que estimar  $p$  parámetros de movimiento (es decir,  $p = 1$  para posición,  $p = 2$  para posición y velocidad,  $p = 3$  para posición, velocidad y aceleración), es posible deducir que:

$$\sigma_N = \frac{p}{\sqrt{N}} \sigma_1$$

Por tanto, la exactitud se degrada continuamente a medida que se estiman más parámetros. Éstas son las exactitudes que se pueden lograr por medio de un rastreador (seguidor) de  $p$ -ésimo orden empleando  $N$  mediciones.

## 51.7 EFECTOS DE LA TRAYECTORIA DE PROPAGACIÓN

Los factores principales que influyen en la propagación de ondas de radar son la trayectoria múltiple, refracción o flexión, y la atenuación y dispersión por los gases atmosféricos.

### 51.7.1 Trayectoria múltiple

La trayectoria múltiple se origina cuando un objetivo es "iluminado" por un rayo directo y un rayo reflejado en la superficie, como se ilustra en la figura 51-10. La intensidad del campo eléctrico resultante en el punto  $T$  es función de la diferencia en longitud de trayectoria entre las trayectorias directa e indirecta, el coeficiente de reflexión de la superficie y el patrón direccional y ángulo de dirección de la antena.

Para un patrón de antena (diagrama de directividad) de haz amplio en ángulo de elevación, como en el caso de un radar bidimensional que sólo mide la distancia y el acimut, los efectos del patrón de antena son mínimos. Si existe un desfase  $\Delta$  en la superficie y una diferencia de longitud de trayectoria igual a  $\epsilon$ , y si la magnitud del coeficiente de reflexión es  $\rho$ , entonces la intensidad de campo resultante en  $T$  es proporcional a:

$$\gamma = \sqrt{1 + \rho^2 + 2\rho \cos[\Delta + (2\pi/\lambda)\epsilon]}$$

La energía que regresa al receptor en viaje de ida y vuelta es proporcional a  $\gamma^4$ , de manera que la energía recibida de un objetivo a la distancia  $R$  y para una sección transversal del radar  $\sigma$  es:

$$P_{\text{REC}} = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} \left[ 1 + \rho^2 + 2\rho \cos\left(\Delta + \frac{2\pi}{\lambda} \epsilon\right) \right]^2$$

Como se muestra en la figura 51-10, si  $h_R$  y  $h_T$  son las alturas del radar y el objetivo sobre la superficie de reflexión, y  $d_0$  es su separación a lo largo del plano horizontal, entonces la diferencia de longitud de la trayectoria puede determinarse aproximadamente como:

$$\epsilon = 2 \frac{h_T h_R}{d_0}$$

Si la magnitud del coeficiente de reflexión es 1 y su fase  $\Delta$  es igual a  $180^\circ$ , que es lo normal sobre el mar con la polarización horizontal, demuestra que:

$$\gamma^4 \approx 16 \sin^4 \left( \frac{2\pi}{\lambda} \cdot \frac{h_T h_R}{d_0} \right)$$

Esto duplica la distancia de detección del objetivo, comparada con la que se obtiene con el radar en el espacio libre, para objetivos donde el argumento de la función  $\sin^4$  es como sigue:

$$\frac{2\pi}{\lambda} \cdot \frac{h_T h_R}{d_0} = \left( N + \frac{1}{2} \right) \pi$$

El primer máximo ocurre en  $N = 0$ , y los máximos de detección sucesivos en  $N = 1, 2, 3, \dots$ . Un valor nulo, o respuesta cero, se obtiene en la zona de desvanecimiento de trayectoria múltiple, donde el argumento de la función  $\sin^4$  es un múltiplo de  $\pi$ , de manera que:

$$\frac{2\pi}{\lambda} \cdot \frac{h_T h_R}{d_0} = N\pi$$

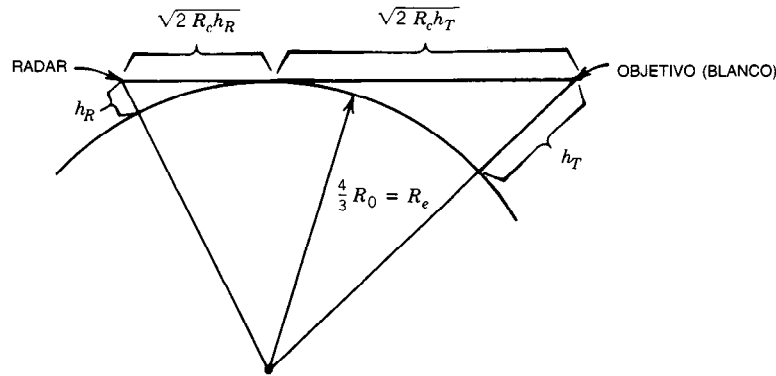


Fig. 51-11. Geometría de la distancia del horizonte del radar.

Por tanto, cuando un objetivo en el aire vuela en una trayectoria radial entrante de altura especificada hacia el radar, éste atravesará alternativamente zonas de desvanecimiento y acentuación en función del tiempo.

El efecto real de la trayectoria múltiple será menos grave que el calculado aquí debido a los siguientes factores:

1. Superficie de dispersión esférica, debido a la forma esférica de la Tierra.
2. Rugosidad de la superficie de dispersión, debida a ondas con altura de onda eficaz (o rms) especificada.
3. Directividad del patrón de la antena del radar, que "ilumina" la superficie del mar con menos intensidad en el caso del rayo de trayectoria múltiple que en el del rayo directo.

### 51.7.2 Refracción

La radiación de los transmisores del radar sufre refracción o flexión cuando las ondas se propagan por la atmósfera. Esto se debe al cambio gradual en el índice de refracción de la atmósfera con la altura creciente, debido a su vez a variaciones en humedad, presión y temperatura atmosféricas. Como resultado, las ondas del radar se propagan cerca del 15 % más allá del horizonte geométrico. El cálculo de estas trayectorias curvas de las ondas del radar es muy complejo y por tanto se determina en forma aproximada suponiendo un radio terrestre de cuatro tercios, como se muestra en la figura 51-11. Así, para una antena de radar montada a una altura  $h_R$  y un objetivo situado a una altura  $h_T$ , la distancia del horizonte del radar está dada por:

$$R_H = \sqrt{2R_e h_R} + \sqrt{2R_e h_T}$$

donde  $R_e$  es el radio terrestre equivalente dado por  $4/3$  por su valor real  $R_0$ . El factor  $4/3$  supone una temperatura y una presión específicas al nivel del mar y cierta disminución de estas cantidades con la altura. Ésta se conoce como atmósfera estándar. También se supone una humedad relativa constante. Cuando se consideran factores atmosféricos variables sobre el globo,

$R_e$  puede ser tan grande como  $2R_0$ , que implica una distancia de horizonte todavía mayor. La fórmula de  $R_H$  no incluye los efectos de trayectoria múltiple que se analizaron antes. El término superrefracción se refiere al caso en el cual los rayos se flexionan hacia abajo más que en una atmósfera estándar. Esta condición suele presentarse cuando una masa de aire se ha calentado, después de pasar sobre una masa de tierra caliente, y fluye sobre un mar relativamente frío. En el mar Mediterráneo, que está casi completamente rodeado por tierra, se observa superrefracción el 90 % del tiempo en primavera y verano.

Si los rayos del radar se flexionan hacia abajo aún más, se pueden reflejar en forma repetitiva y volver a curvarse hacia abajo, como se ilustra en la figura 51-12. Este fenómeno se conoce como propagación por conductos atmosféricos, porque los rayos parecen estar confinados a un conducto o tubo en forma muy parecida a como ocurre en una guía de ondas. Con este fenómeno, una transmisión de radar y su eco pueden propagarse a muy largas distancias (es decir, muchas veces el intervalo de distancia ambigua del radar) antes de que escapen del ducto. En estas circunstancias una imagen PPI del radar puede ser muy confusa.

### 51.7.3 Atenuación y dispersión

En una atmósfera despejada, la energía del radar disminuye debido a la absorción por el oxígeno y vapor de agua. Parte de la energía que incide en las moléculas es absorbida y se pierde como calor. Éste es principalmente un efecto de mecánica cuántica en el cual el estado molecular pasa de un nivel de energía a otro superior. La niebla, nubes, nieve, granizo y lluvia absorben y dispersan también energía incidente del radar. La cantidad de radiación dispersada depende de las propiedades dieléctricas de los dispersores, la polarización de la radiación incidente y el tamaño relativo del dispersor respecto a la longitud de onda de la radiación. Cuando las partículas son mucho menores que la longitud de onda, la energía es dispersada de manera casi isotrópica (es decir, uniforme en ángulo). Éste es el caso usual de las ondas de radar que interactúan con los componentes atmosféricos. Las nubes y la precipi-

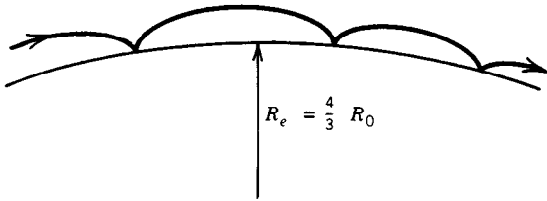


Fig. 51-12. Geometría de la propagación por conductos atmosféricos.

tación pueden generar ecos en el receptor debidos a esta dispersión, y éstos son visibles en la pantalla. Si los ecos son intensos interferirán en la detección normal del objetivo. Si el objetivo se desplaza con una velocidad radial suficiente con respecto a este emborronamiento climático, su presencia se puede determinar por medio del procesamiento Doppler o MTI.

### 51.8 RADAR DOPPLER Y MEDICIONES DE VELOCIDAD

Los radares Doppler tienen la capacidad de detectar objetivos en movimiento en presencia de emborronamiento intenso, tal como el generado por ecos parásitos del mar, la tierra y climáticos. El movimiento del objetivo respecto al transmisor y receptor del radar produce por interferencia un desplazamiento de frecuencia en la señal recibida que no está presente en la señal que regresa del emborronamiento. De este modo, el radar Doppler puede discriminar entre señales de objetivos en movimiento y estacionarios.

Si un objetivo del radar tiene una velocidad radial  $\dot{r}$  con respecto al receptor y transmisor (aquí se supone que están en el mismo sitio), la frecuencia de la señal recibida será:

$$f_R = f_0(1 - 2\dot{r}/C) = f_0 - 2\dot{r}/\lambda$$

donde  $f_0$  = frecuencia del transmisor

$C$  = velocidad de la luz

$\lambda$  = longitud de onda =  $f_0 C$

La cantidad que se suma a la frecuencia transmitida  $f_0$  se conoce como frecuencia Doppler  $f_d$ . Así:

$$f_d = 2\dot{r}/\lambda$$

Si el objetivo tiene una componente de movimiento de salida (llegada) radial, entonces la frecuencia Doppler es negativa (positiva). Por cada nudo de movimiento radial del objetivo a  $f_0 = 10$  GHz, la frecuencia de la señal recibida cambia en 34.29 Hz.

La velocidad radial del objetivo se puede medir en forma casi instantánea con un radar de vigilancia si se utiliza un procesador de señales, como el de la figura 51-13. En este procesador, llamado procesador Doppler de pulsos, se analiza el espectro de la señal. Se transmite un tren coherente en fase de  $N$  pulsos con un periodo entre pulsos constante  $T$  (es decir, PRF constante). La señal recibida, que ha inducido una modulación de fase  $\phi(t)$  debido al movimiento del objetivo, se combina con ruido interferente más el emborronamiento, que se indica como  $n(t) + C(t)$ . La señal de entrada se mezcla con dos señales de referencia en cuadratura de fase (o sea, las dos señales de referencia están defasadas  $90^\circ$  eléctricos entre sí) en canales separados. Las componentes de baja frecuencia de esta operación de mezclado se denominan  $I$  y  $Q$ , las componentes en fase y en cuadratura de la señal respectivamente. Después se realiza una operación MTI en  $I$  y  $Q$  para reducir los efectos del emborronamiento de baja velocidad. La operación MTI más simple de este tipo se obtiene restando dos muestras sucesivas de  $I$  y  $Q$  a la misma distancia, de manera que las salidas del indicador del objetivo en movimiento (MTI) son:

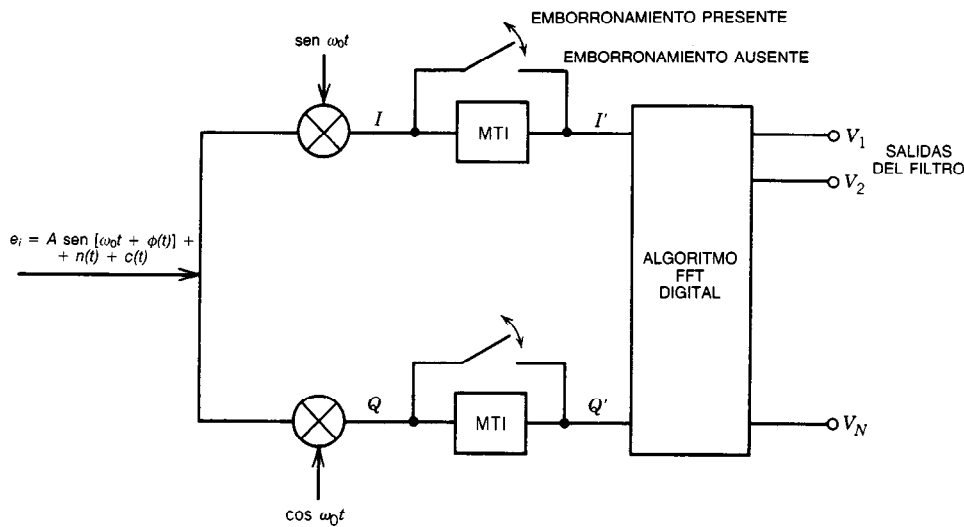


Fig. 51-13. Procesamiento por filtro Doppler. FFT = transformada rápida de Fourier; MTI = indicador del objetivo en movimiento.

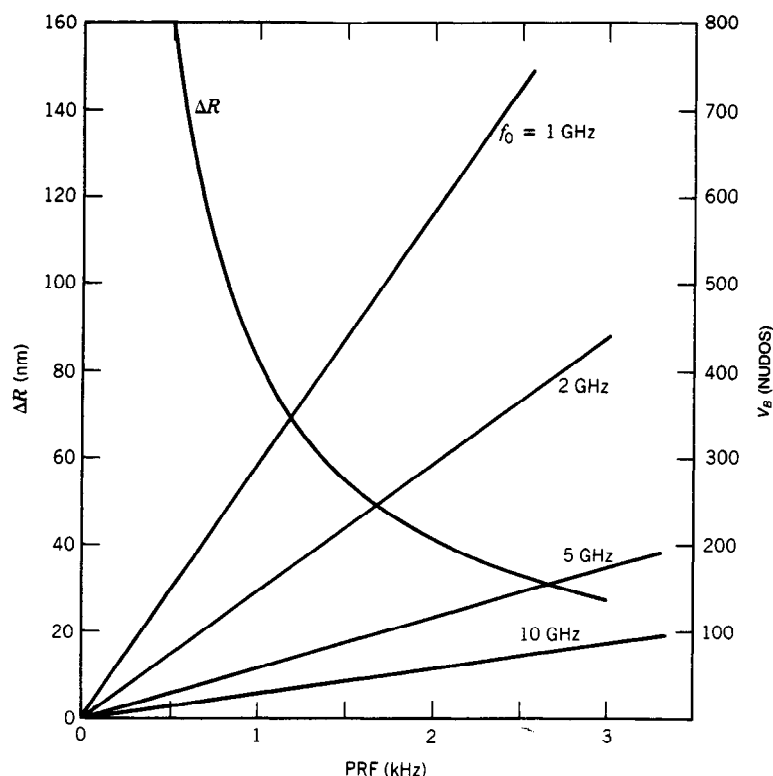


Fig. 51-14. Primera velocidad ciega del radar MTI o Doppler de pulsos,  $V_B$ . También se ilustra la máxima distancia inequívoca,  $\Delta R$ , en nm.  $f_0$  = frecuencia del transmisor.

$$I'(t) = I(t) - I(t - T)$$

$$Q'(t) = Q(t) - Q(t - T)$$

donde  $T$  es el periodo entre pulsos. Las operaciones MTI de este tipo son sensibles a la velocidad radial del objetivo. De hecho, las respuestas del objetivo se cancelan si  $f_d T = M$ , donde  $M$  es un entero. Cuando  $M = 1$ , el valor de  $f_d$  que corresponde a  $1/T$  se conoce como primera velocidad ciega.

Este valor de  $f_d$ , que se define como  $V_B$  en nudos, se grafica en la figura 51-14 como función de  $PRF = 1/T$  y de la frecuencia transmitida  $f_0$ . La relación está dada como:

$$V_B = 291.58 \left[ \frac{PRF \text{ (kHz)}}{f_0 \text{ (GHz)}} \right] \text{ nudos}$$

Por tanto, para  $PRF = 1$  kHz y  $f_0 = 1$  GHz (es decir,  $10^9$  Hz), el valor de  $V_B$  es 291.58 nudos. Así, los objetivos con una velocidad de cambio de la distancia de 291.58 nudos tendrán una respuesta cero fuera del MTI. También en la figura se muestra la distancia inequívoca del radar  $\Delta R$  para una onda PRF fija. Esta relación está dada por  $\Delta R = 81/PRF$  (kHz) nm.

Por último, las componentes  $I'$  y  $Q'$  entran en un procesador FFT digital que analiza el espectro de la señal en un banco de filtros Doppler con la cobertura

como se muestra en la figura 51-15. Este banco de filtros resulta cuando se emplea un FFT de ocho puntos. Las respuestas del filtro, como se muestra, son periódicas con periodo igual a  $V_B$ , la primera velocidad ciega. Por tanto, si  $V_B = 400$  nudos, cada filtro se encarga de un intervalo Doppler de  $400/8 = 50$  nudos. Un objetivo de 200 nudos excitará el filtro central. Objetivos con una velocidad de cambio de distancia de 600, 1 000 o 1 400 nudos también excitarán este filtro central. Esta ambigüedad Doppler se puede resolver mediante el seguimiento o rastreo de la distancia del objetivo con el tiempo o mediante la transmisión de otros lotes de pulsos a una PRF diferente. (Esta técnica elimina asimismo posibles ambigüedades de distancia.) En este último caso se excitará un filtro distinto. Con parámetros de radar adecuados la velocidad de cambio de la distancia del objetivo se puede determinar en forma instantánea con una exactitud mejor que 5 nudos. Con la aplicación de este método en sitios separados, se puede determinar el vector velocidad del objetivo.

Si se detectan objetivos en regiones donde no hay emborronamiento, la operación MTI se puede pasar por alto mediante el cierre de los interruptores de la figura 51-13. Ahora es posible detectar objetivos en múltiplos de la velocidad ciega.

El procesador Doppler de pulsos se puede utilizar con un radar de vigilancia por exploración si existe el

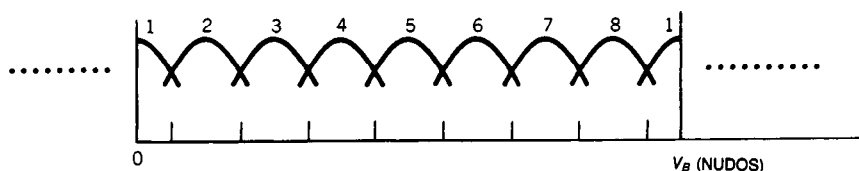


Fig. 51-15. Banco de filtros Doppler.  $V_B$  = primera velocidad ciega.

tiempo de "iluminación" suficiente sobre el objetivo, o con un radar de rastreo (seguimiento) automático (de persecución). También se pueden utilizar otras modulaciones, tales como la onda continua (CW) u onda continua interrumpida (ICW, CW encendida y apagada con alta velocidad de repetición y factor de trabajo elevados). El espectro de la señal devuelta se puede analizar en un banco de filtros Doppler digital o analógico. Se requiere la modulación de la frecuencia de barrido para determinar la distancia del objetivo en transmisiones CW o ICW. Esto se puede realizar mediante la modulación lineal o sinusoidal de la frecuencia del transmisor.

La determinación de la distancia de esta forma es por lo general menos exacta que la que se puede lograr con un procesador Doppler de pulsos.

### 51.9 TÉCNICAS AVANZADAS (ONDAS MILIMÉTRICAS)

El desarrollo de sistemas de radar de ondas milimétricas acrecentará la capacidad de resolución de los sistemas de navegación en el futuro. Con el uso de frecuencias mayores de 30 GHz, la resolución angular mejorará cuando menos tres veces para aberturas de antena

de tamaño similar. Los anchos de pulso más estrechos mejorarán asimismo la resolución en la dimensión de la distancia. Estos sistemas tendrán unas capacidades de alcance limitadas debido a las características de propagación atmosférica en estas frecuencias, de manera que sólo serán útiles en aplicaciones de corta distancia. Se requerirá un sistema de doble frecuencia (conservando la banda  $X$  actual) si se requiere cobertura de larga distancia.

En el futuro cercano comenzarán a producirse sistemas útiles como auxiliares para la evasión de colisiones<sup>5</sup>. La medición del diagrama característico de señales reflejadas hará posible la identificación y el seguimiento automático de objetivos, con el rechazo del emborronamiento y ecos no deseados. Con un sistema de ondas milimétricas resulta fácil prever el uso de técnicas de comparación de mapas en una computadora para cotejar una imagen generada por radar con mapas almacenados a fin de trazar una posición en el sistema de coordenadas del mapa.

También se pueden prever sistemas de telemetría por láser como auxiliares en el procedimiento de entrada en dársena. La posición y las velocidades medidas relativas se pueden emplear para ubicar automáticamente el vehículo transmisor en este delicado procedimiento de corta distancia.



## CAPÍTULO 52

# Auxiliares de la navegación

### **Frederick B. Pogust**

Eaton Corporation  
Farmingdale, Nueva York

### **Stephen C. Martin**

Lockheed Missiles and Space Company  
Palo Alto, California

### **Thomas J. Cutler**

United States Naval Academy  
Annapolis, Maryland

#### **52.1 Normas y convenios internacionales de navegación aérea**

- 52.1.1 Historia
- 52.1.2 Normalización internacional
- 52.1.3 Normalización en Estados Unidos
- 52.1.4 El proceso de estandarización

#### **52.2 Auxiliares de la navegación aérea**

- 52.2.1 Radioayuda para la navegación de aeronaves

- 52.2.2 Radioayuda para la navegación de corta y mediana distancia

- 52.2.3 Sistemas de radionavegación de larga distancia de aeronaves

- 52.2.4 Equipo de detección meteorológica

#### **52.3 Auxiliares de la navegación marítima**

- 52.3.1 Introducción

- 52.3.2 Auxiliares electrónicos de la navegación



## 52.1 NORMAS Y CONVENIOS INTERNACIONALES DE NAVEGACIÓN AÉREA

*Frederick B. Pogust*

### 52.1.1 Historia

En los primeros días de la aviación, el contacto visual con tierra era una necesidad; y la navegación se realizaba mediante técnicas más similares a las asociadas con el automóvil que con el barco. La intervención y regulación del gobierno eran igualmente primitivas, y desde luego era innecesaria la coordinación internacional.

La evolución del complejo sistema de rutas, normas, convenios y procedimientos locales, nacionales e internacionales que rige actualmente la navegación aérea se inició con la necesidad de volar en condiciones de oscuridad visual.

Ciudades, construcciones rurales, carreteras y vías de ferrocarril (la brújula de hierro) ya no eran indicaciones adecuadas para el aviador. Durante la década de 1920, cuando la aviación se hizo un negocio, ciertos sectores de la industria y el gobierno reconocieron la urgencia de poner orden a lo que se estaba convirtiendo rápidamente en una situación caótica y peligrosa en el aire. En 1926 el Congreso de Estados Unidos redactó la primera Air Commerce Act (ley de comercio aéreo) y el gobierno federal emprendió actividades que, hasta donde concierne a la navegación, constituyen una sociedad entre aquellos que vuelan y los responsables de fomentar una actividad de vuelos en forma segura y eficiente.

Los sistemas de navegación que se emplean ahora en el comercio aéreo son supervivientes de cientos de conceptos que se han propuesto, desarrollado, probado y empleado para cumplir el doble objetivo del vuelo guiado: la necesidad del piloto de contar con información que le asegure que se encuentra en la ruta deseada entre los puntos A y B, y la necesidad que tienen las autoridades de control del tráfico aéreo de contar con información para asegurarse de que el avión Y está situado a la distancia suficiente del avión Z para evitar una colisión. Son estos dos objetivos relacionados pero independientes los que han llevado a la organización de la mayor parte de los vuelos comerciales en un sistema de rutas aéreas señalizadas y deslindadas.

Las primeras rutas aéreas consistían en marcas o indicaciones en tierra a lo largo de la ruta. Con el tiempo se agregaron faros giratorios que emitían un haz luminoso. Hacia 1928, el Department of Commerce operaba las primeras estaciones de radiofaro direccional que generaban cuatro cursos acústicos en la ruta aérea de Nueva York a Cleveland. Para emplear esta nueva forma de radionavegación, el avión tenía que contar con el equipo de recepción adecuado. El gobierno emitía una señal estandarizada desde tierra, y el usuario necesitaba tener un equipo que pudiera interpretar dicha señal. Entonces nació el requisito de normas y convenios.

La institucionalización de los sistemas que se emplean en la navegación aérea avanzó a pasos acelerados en toda la década de 1930. En 1935 la industria formó un grupo coordinador llamado Radio Technical Committee for Aeronautics (RTCA), que desempeñaría un cometido preponderante en la definición de los sistemas que se estandarizarían. En 1938 el Congreso de Estados Unidos redactó una nueva Civil Aeronautics Act (ley de aeronáutica civil) que creaba la Civil Aeronautics Authority (CAA) en el Department of Commerce. La CAA dirigió la evolución de las normas de aeronáutica civil desde la posguerra hasta el sistema actual. La Segunda Guerra Mundial tuvo los efectos contradictorios de poner fin a cualquier variación en los sistemas civiles locales y de acelerar considerablemente el avance de la tecnología que por último los sustituiría.

Para el inicio de la década de 1940, el comercio aéreo realizado por líneas aéreas comerciales se basaba casi en su totalidad en el DC-3. Estados Unidos y otros países desarrollados del mundo estaban enlazados por vías aéreas tendidas a lo largo de las rutas de los radiofaros direccionales. Los aviones de esta era estaban equipados con radios transmisoras de habla para establecer comunicaciones de aire a tierra y con radiogoniómetros que orientaban automáticamente sus antenas hacia una estación sintonizada, proporcionando una lectura de radiobrújula (brújula radiogoniométrica). Estos receptores de mediana y alta frecuencia podían sintonizar asimismo un radiofaro. El vuelo por una vía aérea se realizaba guiando el avión hasta alcanzar un tono estable en los audífonos. Esto indicaba la alineación con uno de los cuatro haces o sectores de rumbo del radiofaro. Una desviación respecto de la ruta aérea daría por resultado una "A" o "N" codificadas en clave Morse en los audífonos del piloto.

El comercio aéreo de ultramar o internacional estaba confinado a un puñado de hidroaviones. Éstos seguían su trayecto con la ayuda de navegantes profesionales, quienes tomaban puntos fijos en el cielo con sextantes manuales, como lo hicieran los marineros durante cientos de años.

La CAA proporcionó el sistema de radiofaros direccionales, radiofaros no direccionales y radiobalizas de abanico, que eran lo usual en esa era. También tenía nuevos sistemas en desarrollo:

1. Radiofaro omnidireccional óptico (VOR), que indicaba un curso seleccionado cualquiera hacia la estación del radiofaro o desde ella.
2. Equipo radiotelemétrico (DME), que indicaba la distancia real a una estación en tierra.
3. Sistema de aterrizaje por instrumentos (ILS), que constaba de un localizador que indicaba la alineación exacta con una pista de despegue y aterrizaje, una trayectoria de planeo que indicaba la velocidad de descenso correcta, y radiobalizas marcadoras que indicaban la distancia a la pista.

Estos sistemas se analizan con detalle en la sección 52.2.

### 52.1.2 Normalización internacional

El término de la guerra estableció la plataforma de transformación de todo el sistema aéreo comercial. Los aviones ahora podían volar sin dificultad sobre los océanos y de un país a otro. Se había perfeccionado la tecnología necesaria para guiarlos, pero muchos de los sistemas que resultaron de esta tecnología no eran compatibles entre sí. Las pocas naciones tecnológicamente avanzadas diferían en las técnicas que mejor se adecuaban a la nueva era. En 1947 se convino la primera reunión de una agencia internacional que trataría estos asuntos, y pronto la International Civil Aviation Organization (ICAO), agencia especializada de las Naciones Unidas, se convirtió en una institución permanente de comercio aéreo con personal y oficinas centrales en Montreal, Quebec. Actualmente, la ICAO tiene entre sus responsabilidades principales la ratificación de todas las normas y convenios de aviación internacionales.

La función de la ICAO es asegurar que las naciones que participan en el comercio aéreo internacional ofrezcan servicios estándares que puedan ser utilizados por aeronaves equipadas en forma adecuada. Esto significa que los aviones pueden volar de un país a otro con un complemento de equipos para realizar los procedimientos requeridos. La ICAO no tiene una función oficial en los vuelos nacionales, pese a que la presión para conformar los sistemas nacionales a las normas de la ICAO es evidente.

Una excepción interesante a esta norma es la Unión Soviética, que equipa los aviones de Aeroflot en comercio internacional con instrumentos estándares de la ICAO y ha instalado auxiliares de la navegación de este organismo en el corto número de aeropuertos de acceso internacional que utiliza. Sin embargo, todos los sistemas internos fueron instalados antes de que la Unión Soviética se uniera a la ICAO y éstos difieren totalmente de las normas. Ahora se están haciendo esfuerzos por convertir los sistemas internos a las normas de la ICAO.

La documentación funcional de este organismo para lograr su objetivo de estandarización son las *International Standards and Recommended Practices (SARPS)*, *Aeronautical Telecommunication*, anexo 10 a la *Convention on International Civil Aviation* (convenio sobre aviación civil internacional). El volumen 1 del anexo 10 incluye la primera parte, *Equipment and Systems* (equipo y sistemas) y la segunda parte, *Radio Frequencies* (radiofrecuencias). El volumen 3 abarca el procedimiento de comunicaciones. Estos volúmenes se han actualizado periódicamente desde su aplicación por vez primera en 1950. En la actualidad el volumen 1 está en su tercera edición. Cuando menos 63 emmendas se han vuelto aplicables desde su publicación original. Se incluyen especificaciones de los siguientes auxiliares de la radionavegación:

1. Sistema de aterrizaje por instrumentos (ILS).
2. Sistemas de radar de precisión para aterrizaje (PAR).
3. Radiofaro omnidireccional (VOR) VHF.

4. Radiofaro no direccional (NDB).
5. Equipo radiotelemétrico (DME) VHF.
6. Radiobalizas marcadoras de curso VHF (75 MHz).
7. Características del sistema de visualización.
8. Radar secundario de vigilancia (SSR).
9. Sistemas receptores ADF (radiogoniómetros automáticos) de aviones.
10. Sistema LORAN-A.
11. Sistema de aterrizaje por microondas (guiaje angular) (MLS).

Éstos son los sistemas estandarizados internacionales actuales. (El DME de precisión que se utilizará con el MLS está en proceso de ser estandarizado.) Además, otros sistemas estándares ajenos a la ICAO se utilizan para la navegación aérea en todo el mundo, incluyendo Estados Unidos. Existe una estructura de normalización totalmente independiente promovida por la Federal Aviation Administration (FAA) y un sistema de reglamentos obligatorios, las Federal Air Regulations (FAR), que también definen los sistemas de navegación que se pueden aplicar. En general, donde se superponen las normas del ICAO y de la FAA, están en completo acuerdo.

### 52.1.3 Normalización en Estados Unidos

En Estados Unidos, la FAA se encarga del desarrollo de sistemas de navegación aérea y de control del tráfico aéreo. Los militares operan sus propios componentes de estos sistemas de conformidad con las normas de la FAA. Cuando la CAA era la autoridad en aviación, no tenía jurisdicción sobre los sistemas militares y existía divergencia. Una disputa entre las agencias civiles y militares sobre la forma del sistema de navegación de corta distancia para uso nacional fue uno de los elementos para la desaparición de la CAA y la formación de la FAA (Federal Aviation Authority) en la Federal Aviation Act (ley de aviación federal) de 1958.

#### Navegación de corta distancia

La manifiesta renuncia del Congreso a dotar de fondos a sistemas civiles y militares independientes, así como los posibles riesgos de una política incoordinada como ésta, llevaron al Special Committee RTCA 31 a realizar un estudio que recomendaba un sistema civil y militar común de navegación y control del tráfico aéreo. La milicia propuso que su sistema TACAN (navegación aérea táctica) fuese adoptado como sistema de navegación común, mientras que la CAA había estado procediendo a la implementación del VOR y su versión de un DME. El DME civil y la TACAN militar, que incluye una señal de tipo VOR y un DME, ocupaban la misma banda en el radioespectro. La resolución de este conflicto dio por resultado el sistema VORTAC. Una estación VORTAC, que es el auxiliar de la navegación de corta distancia estándar operado por la FAA, incluye un sistema de acimut VOR estándar VHF con las normas de la ICAO (usado principalmente por aviones civiles), un sistema de acimut estándar

dar ajeno a la ICAO de banda L (UHF) (utilizado principalmente por aviones militares), y de un DME estándar de banda con las normas de la ICAO (que usan ambos). Pese a que en la práctica sólo unos cuantos aviones civiles (o quizá ninguno) están equipados para volar con el sistema de acimut de banda L, nada les impide hacerlo en el espacio aéreo de Estados Unidos. El inhibidor es el costo.

### Sistemas de aterrizaje

La FAA conserva asimismo un sistema de auxiliares de aproximación y aterrizaje, llamado sistema de aterrizaje por instrumentos o instrumental (ILS), en más de 1 000 pistas de despegue y aterrizaje en aeropuertos para uso público civil. Este sistema, pese a ser una norma de la ICAO garantizada por lo menos hasta 1995, está programado para ser sustituido por un nuevo sistema estándar de la ICAO, llamado sistema de aterrizaje por microondas (MLS), en años futuros. Muchos campos aéreos militares de Estados Unidos y otros países no cuentan con ILS, sino que emplean una forma de sistema de guiaje por radiotelefonía mediante un radar de precisión para aterrizaje (PAR) estándar de la ICAO, que es un elemento de un sistema de aproximación controlado desde tierra (GCA). La FAA en alguna ocasión operó con PAR, pero ya no lo hace. No obstante, en algunos aeropuertos de otros países se pueden encontrar PAR. Los sistemas de aterrizaje de aeronaves se analizan en el capítulo 53.

### Sistemas de radionavegación de larga distancia (LORAN)

En el campo de la radionavegación de larga distancia, la situación es aún más complicada y confusa. Los sistemas de radionavegación de larga distancia suelen utilizarse en vuelos sobre el mar. Esto se debe a que los sistemas de larga distancia utilizan frecuencias en las regiones más bajas del radioespectro, que se propagan más lejos (pero no con la misma eficacia sobre tierra que sobre agua) que las frecuencias superiores de línea recta utilizadas en el VORTAC. En virtud de que los Estados Unidos continentales están cubiertos por 800 sistemas VORTAC, rara vez se requieren sistemas de larga distancia más complejos. Sin embargo, los operadores de helicópteros, que a veces vuelan demasiado bajo para recibir cobertura continua de VORTAC, están interesados en que la FAA estandarice un sistema de baja frecuencia para su uso exclusivo.

El sistema de radionavegación de larga distancia más común que se generó durante la guerra fue el LORAN-A, que es una norma de la ICAO. CONSUL es otro sistema que fue favorecido por los ingleses. Estos sistemas no proporcionaban cobertura a nivel mundial, ni eran tan exactos como se deseaba.

La U.S. Navy and Coast Guard necesitaba un sistema que tuviera una cobertura global. Esta dependencia creó y ha instalado el sistema OMEGA. Este sistema, pese a carecer de exactitud en muchas funciones de navegación aérea (los barcos también utilizan los sistemas de radionavegación de baja frecuencia de larga distancia, como se analiza en la sección 52.3), ofrece cobertura a nivel mundial sólo desde ocho estaciones.

Una versión más exacta de LORAN, a una frecuencia más baja, llamada LORAN-C, se ha difundido ampliamente para la cobertura de las rutas transoceánicas con mayor tráfico. La mayor parte de las estaciones de LORAN-A se han retirado de servicio.

La ICAO no ha estandarizado OMEGA ni LORAN-C. En cambio, han predominado la exactitud y confiabilidad de los llamados sistemas de navegación independientes (el radar Doppler y la navegación inercial) en el uso reciente de líneas aéreas transoceánicas. Esto ha propiciado el retiro de la mayoría de los navegantes humanos. Los sistemas independientes o autónomos no requieren estaciones terrenas y proporcionan la exactitud suficiente para llevar aviones que cruzan el océano directamente a los sistemas de navegación de corta distancia cuando se encuentren dentro de la línea de visión de tierra.

El resultado de esta abundancia de sistemas y técnicas es que no se ha actualizado la estandarización ICAO inicial de sistemas de larga distancia; no ha habido necesidad de hacerlo. Desde el punto de vista de la ICAO, los sistemas autónomos no necesitan conformarse a rígidas especificaciones de compatibilidad (no hay nada con lo cual hacerlos compatibles) mientras sean lo suficientemente confiables para garantizar que los aviones en vuelos internacionales puedan informar su posición con la exactitud que se requiere.

### Participación de la FAA en la reglamentación

El cometido de la FAA en la especificación y regulación de sistemas de navegación difiere del de la ICAO en que:

1. La FAA ofrece un servicio directo subvencionado por el gobierno.
2. La FAA se encarga de garantizar que los aviones sean operados de manera segura dentro del territorio de Estados Unidos.

Por tanto, la FAA debe asegurarse de que, hasta donde ofrece el servicio, limite su gasto sólo a aquellos sistemas que considere necesarios para apoyar a la comunidad de la aviación. Asimismo, ningún sistema de navegación empleado por una aeronave deberá en absoluto representar un riesgo. La FAA actualmente opera sólo el sistema VORTAC y los sistemas ILS en aeropuertos civiles. Los sistemas de radionavegación de larga distancia, es decir, OMEGA y LORAN-C, son operados por la marina y la guardia costera. Los propietarios de aeronaves pueden adquirir en el comercio el equipo necesario para emplear estos últimos, y en realidad esto es muy común en la aviación comercial. La FAA tiene la responsabilidad de regular la calidad de estos productos para garantizar que están en condiciones de volar y certificar que son adecuados para el apoyo del sistema de control de tráfico aéreo.

### Planes para el futuro

El Department of Transportation de Estados Unidos, del cual forman parte la FAA y el servicio de la guardia costera, ha sido nominado por el Congreso para ser el coordinador de una política de navegación nacional.

La navegación aérea, puesto que es la que tiene mayor demanda, tiende a dominar esta política. Los estudios actuales que se realizan para esta política están centrados en el efecto último de la aparición del sistema de navegación por satélite, llamado sistema global de localización (GPS)/NAVSTAR (cap. 54).

GPS/NAVSTAR es un proyecto del Department of Defense de Estados Unidos que promete una cobertura a nivel mundial altamente exacta de cualquier vehículo que esté adecuadamente equipado para recibir las señales cuando menos de cuatro satélites de una galaxia de 18 o 24 satélites. La implicación de un sistema de este tipo es que podría reemplazar a OMEGA, LORAN-C y VORTAC, que actualmente son sostenidos por el gobierno de Estados Unidos, así como a otros sistemas de radionavegación de larga y corta distancia. Sin embargo, para que suceda esto tendría que convertirse asimismo en una norma de la ICAO. El proceso por el cual un sistema militar privado de Estados Unidos podría convertirse en una norma internacional civil no es evidente. La internacionalización futura de un sistema de navegación por satélite tiene muchas ventajas técnicas, aun si los problemas políticos limitan su implementación.

#### 52.1.4 El proceso de estandarización

##### Proceso de la ICAO

El proceso mediante el cual un sistema se convierte en una norma internacional no ha sido estandarizado. En 1950, cuando se dio inicio al anexo 10, Estados Unidos propuso la estandarización de los sistemas que estaban implementándose en ese momento en su territorio. Existía cierta controversia, principalmente debido a que Inglaterra propuso que su sistema de navegación DECCA fuese el de corta distancia. Finalmente los sistemas básicos de Estados Unidos fueron los adoptados.

La mayor parte de las enmiendas que se han hecho en el transcurso de los años han sido el resultado de la experiencia en Estados Unidos, donde el volumen de tráfico aéreo es el más elevado y donde el congestionamiento impone las más grandes demandas a los sistemas de navegación aérea. Por ello, cuando se redactó y ratificó el SARPS, solía reflejar los resultados de la extensa experiencia de la vida real.

La adopción de un nuevo sistema, o el abandono de uno antiguo, es un proceso laborioso y prolongado. Pueden participar más de 130 naciones, y cualquier modificación puede tener graves consecuencias financieras tanto para las naciones grandes como para otras pequeñas, subdesarrolladas. La ICAO forma grupos técnicos especiales para estudiar todas las implicaciones de una variación, y cualquier duda persistente tiende a prolongar el proceso. Cuando establece una norma, este organismo garantiza que lo seguirá siendo en un periodo prolongado para garantizar a las naciones que el equipo que adquieran no se volverá rápidamente obsoleto y para asegurar a los propietarios de aviones que pueden invertir sin riesgo en equipo aéreo.

El costo y la dificultad de hacer adiciones a la lista de sistemas son ilustrados por el proceso a través del cual

el MLS se ha convertido recientemente en una norma de la ICAO. Estados Unidos ya no puede estar seguro de que un voto en la asamblea de la ICAO dará apoyo a su postura. Por otra parte, ni siquiera dicho país puede darse el lujo de crear un nuevo sistema, instalarlo para uso nacional, manejar grandes volúmenes de datos que demuestren su idoneidad y después arriesgarse a un posible rechazo por parte de la ICAO. Por tanto, el desarrollo del MLS se emprendió como parte de un esfuerzo internacional. El vehículo que eligió la FAA para encabezar la creación de un nuevo sistema de aterrizaje fue la RTCA.

La RTCA es un grupo coordinador conjunto del gobierno y la industria que acoge y apoya la participación internacional. En 1963 se fundó en Europa una organización similar, conocida como European Organization for Civil Aviation Electronics (EUROCAE). Las dos organizaciones trabajan en paralelo y coordinan efectivamente sus esfuerzos. Su función es proporcionar consultoría técnica a los gobiernos que deben tomar las decisiones de implementación. El procedimiento implica la identificación de un problema técnico u operacional y después la formación de un comité para estudiar este asunto. Su informe es básicamente una recomendación que, como tal, no tiene fuerza legal.

La función de la RTCA en la creación de un nuevo sistema de aterrizaje fue estrictamente consultiva. Después de tres años de deliberación, un comité especial recomendó que se desarrollara más a fondo uno de los dos tipos de sistema. A esta conclusión se llegó después de estudiar los requisitos operacionales así como las alternativas técnicas.

La ICAO reconoció también el interés que existe en reemplazar el ILS e instituyó sus propios estudios en el All-Weather Operations Panel (AWOP), dedicado a la documentación de requisitos operacionales. La pertenencia a la ICAO está limitada a los gobiernos. Sus grupos de trabajo están constituidos por individuos que han sido designados por sus gobiernos. Muchos de estos mismos individuos participan en actividades de RTCA y EUROCAE, ya que esos organismos abren sus puertas a miembros del gobierno y de la industria.

La propuesta real de corregir el SARPS sólo puede ser hecha por un gobierno miembro de la ICAO. Un total de cinco naciones propusieron sus diseños de sistemas de aterrizaje específicos al grupo de navegación aérea de la ICAO como sustitutos de ILS. Se designó un consejo de doce miembros para revisar estos sistemas y recomendar uno de ellos a la Asamblea de la ICAO. Este organismo tiene la representación de todos los Estados miembros.

Francia, Gran Bretaña, Alemania, Australia y Estados Unidos hicieron propuestas de MLS. Los dos últimos países propusieron el mismo sistema. El proceso de toma de decisiones fue tanto técnico como político. La Unión Soviética dio su apoyo a la propuesta de Estados Unidos y Australia, y, después de algún ajuste de sus conceptos, Alemania se unió a este grupo y se optó por el concepto de haz de barrido con referencia de tiempo (TRSB).

El proceso de ratificación formal y bosquejo del lenguaje exacto del SARPS llevó cerca de tres años. El

MLS ha sido incorporado por enmienda al anexo 10 del convenio. La implementación del sistema requerirá muchos años más. Se garantiza a los usuarios que el ILS seguirá siendo la norma en los aeropuertos internacionales cuando menos hasta 1995, y no se requerirá la disposición de señales MLS ni de equipo MLS en aeronaves por muchos años más.

### Proceso en Estados Unidos

Pese a que cuando se debate un asunto importante como el MLS la función de la RTCA recibe alguna atención del público, el trabajo real de la organización se asigna a sus comités especiales. En un momento cualquiera, están activos hasta quince comités. Desde 1935 ha habido más de 150 comités especiales. Del trabajo que ellos realizan, el que atañe más directamente a las normas que se aplican a la navegación es la publicación de las normas operacionales mínimas de funcionamiento (MOPS) para los diversos sistemas y equipos. Estos documentos se emiten como pauta para los constructores de equipos de aviación. A menudo se incorporan totalmente o por referencia en regulaciones de la FAA. Los documentos oficiales de la RTCA, que se publican con cubiertas verdes distintivas, representan el mejor consenso posible de requisitos técnicos que se pueden resolver entre el gobierno, fabricantes de equipo y usuarios.

La FAA implementa y vigila el cumplimiento de las normas de navegación aérea nacional en Estados Unidos por varios procesos. Primero, se acostumbra que las porciones de tierra de un sistema de navegación sean propiedad del gobierno federal. Si bien es posible la comisión de un sistema auxiliar no federal, este sistema debe coincidir con las normas de la FAA si va a estar abierto al uso público. Varios estados o localidades han comprado e instalado, de vez en cuando, auxiliares de navegación propios, pero siempre siguiendo las normas de la FAA. También es posible operar un auxiliar de navegación privado en Estados Unidos siempre que el público no tenga acceso a él y que la FAA esté convencida de que no representa un riesgo para el usuario o para el público. La FAA emite un certificado técnico especial para este uso privado.

Un control más sobre el uso de instalaciones de navegación es el ejercido por los requisitos de la FAA de que una aeronave debe recibir servicios de control de tráfico aéreo. En algunas condiciones climatológicas, todas las aeronaves deben operar previa autorización del control de tráfico aéreo, y algunas clases de aeronaves de transporte aéreo deben operar siempre con dicha autorización. La FAA tiene la facultad (y la ejerce) de exigir que estas aeronaves cuenten con instalaciones de navegación adecuadas y aprobadas.

Los sistemas de navegación VOR, DME e ILS son los que más fácilmente cumplen este requisito [también suelen requerirse equipo de comunicaciones y equipo de radiofaro respondedor (baliza de radar) para control del tráfico aéreo]. Sin embargo, se han propuesto otros sistemas como el LORAN-C u OMEGA como posibles calificadores. La FAA estudia estas propuestas sobre una base individual. El uso de sistemas no estándares dentro del espacio aéreo nacional es raro.

Finalmente, la FAA en su carácter de autoridad de certificación autoriza el uso de cada elemento específico de equipo para aeronaves. El diseño y fabricación de unidades para uso aéreo se deben realizar según las normas de la FAA antes de su distribución al público propietario de aeronaves. La FAA emite una orden de norma técnica (TSO) por cada elemento que aprueba. Está prohibido por la reglamentación aérea federal vender o utilizar un dispositivo de navegación para el tráfico aéreo que no haya recibido una TSO. Como la emisión de una TSO está basada en pruebas metódicas, se halla regulada por una estandarización rigurosa. El equipo debe realizar aquella función para la que fue creado.

La evolución de la navegación aérea ha sido lenta y constante desde los primeros vuelos. No existe razón para creer que esta tendencia no continuará. Hasta ahora, la radiolocalización de cuatro cursos (rumbos) que precedió a la estandarización ICAO es el único sistema que ha desaparecido en su totalidad. Es probable que pasen muchas décadas antes de que cualquier otro sistema quede en el olvido.

Por otra parte, las tendencias para el futuro parecen apuntar a sistemas de navegación independientes (autónomos) o sistemas de tierra basados en satélites. Estos, sin duda, traerán consigo métodos nuevos e imaginativos para asegurar normas adecuadas dentro de convenios funcionales.

### LIBROS Y REVISTAS

- *Federal Aviation Regulations*, Parte 37, "Technical Standard Order Authorizations", Department of Transportation, Federal Aviation Administration, Washington, DC.
- *Federal Radio Navigation Plan*, Department of Transportation, Federal Aviation Administration, Washington, DC.
- *International Standards and Recommended Practices, Aeronautical Telecommunications*, Anexo 10, Convention on International Civil Aviation, Volumen I, 3.ª ed. (aumentado).
- R. J. Kent, Jr., "Safe, Separated and Soaring; A History of Federal Civil Aviation Policy, 1961-1972", US Department of Transportation, Federal Aviation Administration, Washington, DC.
- N. A. Komins, "Bonfires to Beacons, Federal Civil Aviation Policy Under the Air Commerce Act, 1926-1938", US Department of Transportation, Federal Aviation Administration, Washington, DC.
- S. I. Rochester, "Takeoff at Mid-Century; Federal Civil Aviation Policy in the Eisenhower Years, 1953-1961", US Department of Transportation, Federal Aviation Administration, Washington, DC.
- J. R. M. Wilson, "Turbulence Aloft, the Civil Aeronautics Administration Amid Wars and Rumors of Wars, 1938-1953", US Department of Transportation, Federal Aviation Administration, Washington, DC.

## 52.2 AUXILIARES DE LA NAVEGACIÓN AÉREA

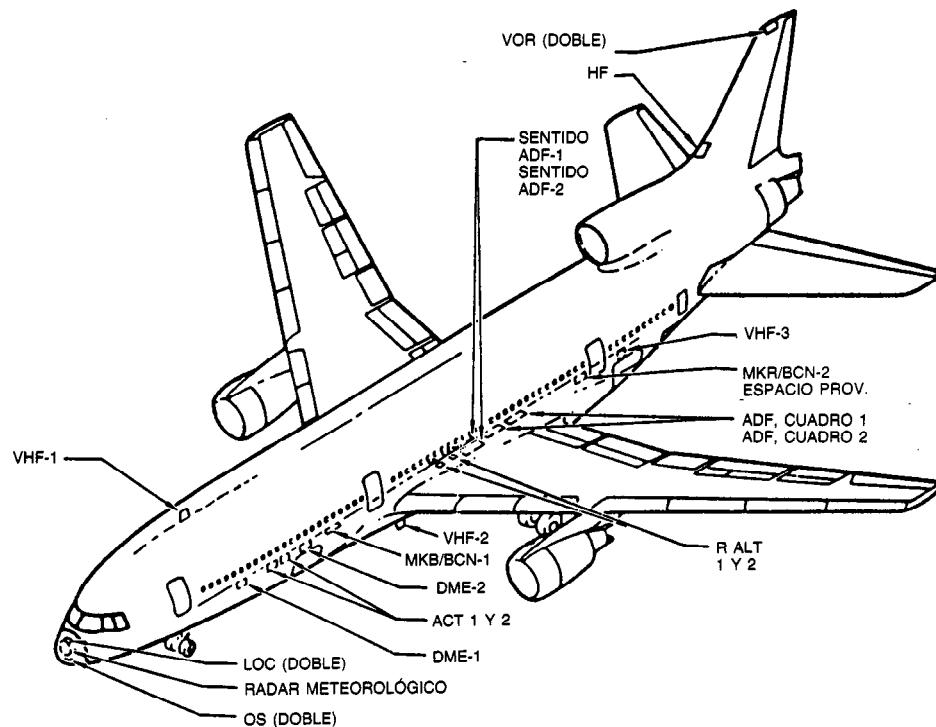
Stephen C. Martin

### 52.2.1 Radioayuda para la navegación de aeronaves

La capacidad de las aeronaves comerciales, militares y privadas de navegar y aterrizar en casi todas las condiciones climatológicas y de visibilidad se considera ahora algo muy natural. Este desempeño se logra a

través del uso de muchas radioayudas distintas, la mayor parte de las cuales son la culminación de más de 40 años de evolución. Para poner un ejemplo, la figura 52-1 ilustra todas las ubicaciones de antenas de una aeronave comercial moderna que se requieren para las variadas funciones de comunicación, navegación, vigilancia y aterrizaje por instrumentos. Éste es el complemento mínimo que se requiere para la navegación continental.

Para vuelos en regiones oceánicas y remotas, se utilizan otras radioayudas de navegación a larga distancia o navegación inercial.



Antena	Uso
VHF (muy alta frecuencia) HF (alta frecuencia)	Comunicaciones
VOR (radiobaliza omnidireccional VHF) ADF, cuadro y sentido (radiogoniómetro automático) DME (equipo radiotelemétrico)	
ATCRBS (sistema de radiobaliza para el informe del control del tráfico aéreo)	Vigilancia del control del tráfico aéreo desde tierra
LOC (localizador) ADF (localizador de brújula) GS (trayectoria de descenso) MKR/BCN (radiobaliza marcadora de 75 MHz) R ALT (altímetro de radar)	

Fig. 52-1. Ubicaciones de antenas en un avión. De Lockheed California Company, con permiso.

Tabla 52-1. Principales sistemas auxiliares para la navegación de corta y mediana distancia

Sistema	Intervalo de frecuencia	Cobertura utilizable	Número de usuarios proyectado		
			Civil E.U.A.	Militar E.U.A.	Todo el mundo
VOR (radiofaro omnidireccional VHF)	108 a 117.95 MHz	Línea de visión	196 000	12 000	270 000
DME (equipo radiotelemétrico)	962 a 1213 MHz	Línea de visión	70 000	13 000	130 000
TACAN (navegación aérea táctica)	962 a 1213 MHz	Línea de visión	— —	13 000	16 000
NDB/ADF (radiofaro no direccional/radiogoniometría automática)	200 a 1650 kHz	50 a 200 nmi	100 000	24 000	170 000
DECCA Radiobaliza marcadora	70 a 130 kHz 75 MHz	≈ 210 nmi < 6 nmi	— — 77 000	— — 12 000	1 000 150 000
SSR (radar secundario de vigilancia)	1030 a 1090 MHz	Línea de visión	180 000	24 000	250 000

Fuente: Federal Radio Navigation Plan<sup>1</sup>.

En las grandes aeronaves de transporte modernas, el manejo de estos diversos sistemas auxiliares de navegación es asistido por computadoras de gobierno de vuelo. Éstas son capaces de seleccionar y sintonizar los diversos sistemas en forma automática y generar una sola visualización navegacional con base en combinaciones de entradas. La más reciente generación de aeronaves presenta estos datos a la tripulación mediante el uso de pantallas CRT (tubo de rayos catódicos) de color de funciones múltiples, integrados con otros datos de la nave, tales como el rumbo, velocidad del aire, y ángulos de balanceo y cabeceo, además de datos del equipo de detección climatológica de a bordo.

Aquí se estudiarán las radioayudas de la navegación en ruta (de corta y larga distancia) y para la detección climatológica. En el capítulo 53 se analizan las radioayudas para el aterrizaje por instrumentos.

#### 52.2.2 Radioayuda para la navegación de corta y mediana distancia

En la tabla 52-1 se resumen algunos sistemas importantes que dan servicio hasta a 200 nmi (millas náuticas) desde la estación de tierra (dependiendo de la altura de la aeronave). De estos sistemas, todos menos el DECCA (sistema británico) y la parte de medición del acimut del TACAN tienen sus parámetros estandarizados por convenio de la ICAO.

El radiofaro tetradireccional de baja frecuencia (ref. 2, cap. 2), primer auxiliar de la navegación que se uti-

lizó ampliamente al inicio de la década de 1920, es ahora obsoleto. La radiobaliza marcadora de VHF, que se creó originalmente para usarse con este sistema, ha sobrevivido para el uso en ruta; pero ahora se utiliza principalmente como parte del sistema de aterrizaje por instrumentos (ILS).

El radar secundario de vigilancia o exploración (SSR), conocido en Estados Unidos como sistema de radiobaliza para el informe de posición del control de tráfico aéreo (ATCRBS), no es un sistema de navegación de aeronaves en sí, sino un auxiliar para la vigilancia desde tierra del tráfico aéreo. En la actualidad se está perfeccionando para que sirva como un enlace de datos para la transferencia de mensajes ATC estandarizados y como base de un esquema automatizado de evasión de colisiones.

#### Sistema de radiofaro omnidireccional de muy alta frecuencia (VOR)

El sistema de radiofaro omnidireccional de muy alta frecuencia (VOR) o de onda métrica es el principal auxiliar de la radionavegación de corta y mediana distancias que utilizan las aeronaves civiles. Fue adoptado originalmente como norma internacional por la ICAO en 1949. Las estaciones de tierra VOR se encuentran por todo el mundo, aunque pocos países, salvo Estados Unidos, han logrado ya casi el 100% de cobertura VOR del espacio aéreo nacional.

El VOR es un auxiliar de la navegación de fuente puntual, que proporciona al usuario una medición del

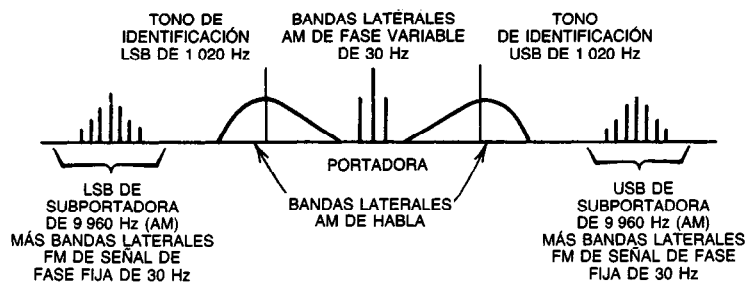


Fig. 52-2. Espectro de una señal VOR (radiofaro omnidireccional VHF) ordinaria. LSB = banda lateral inferior; USB = banda lateral superior.

rumbo acimutal a la estación referido al norte magnético local. A diferencia de los sistemas radiogoniométricos o de radioguía de recalada a bordo, el sistema VOR no produce una lectura como función del rumbo de la aeronave. La estación de tierra radia una onda portadora VHF simultáneamente modulada por dos señales de 30 Hz. Una componente de 30 Hz es de fase fija y se conoce como señal de referencia. La fase de la segunda señal, variable, cambia como función del acimut a la estación. La diferencia de fase entre las dos señales proporciona una medida del acimut a la estación. Como el enfasamiento de las señales se fija de manera que en algún lugar se observe una diferencia de fase cero debido al norte magnético de la estación, la diferencia de fase real ( $0$  a  $360^\circ$ ) que mide una aeronave es igual a su ángulo de acimut a la estación.

En un VOR ordinario, la señal de fase variable se genera aplicando una portadora no modulada a un sistema de antena con diagrama cardioide de acimut que gira a 30 r/s. A partir de esta componente el usuario observa una modulación de amplitud de doble banda lateral (DSB) de la portadora en 30 Hz. La señal de fase de referencia se transmite desde una antena omnidireccional. La portadora se modula primero en amplitud DSB con una subportadora de 9 960 Hz. Esta subportadora se modula en frecuencia en forma sinusoidal a 30 Hz con desviación de cresta de  $\pm 480$  Hz. La fase detectada de esta modulación de 30 Hz constituye la referencia fija.

Superpuestas a la portadora de la señal de referencia están las señales de identificación codificadas en clave Morse y moduladas en amplitud que usan un tono de 1 020 Hz. Además, en algunos sitios se agregan señales de habla con fines de identificación, predicción climatológica o comunicaciones por medio de modulación de amplitud. El espectro compuesto resultante se ilustra en la figura 52-2.

Las frecuencias de operación VOR están en la banda de 108 a 118 MHz. Se han asignado canales para un espaciamiento de 50 kHz. Actualmente sólo se utilizan en realidad canales con frecuencias en incrementos de 100 kHz (p. ej., 112.10 MHz, 114.30 MHz). La distancia de recepción es esencialmente rectilínea, con poco o nada de contaminación por la propagación de ondas ionosféricas. Se utiliza polarización horizontal. En las pruebas iniciales de los sistemas prototipo<sup>3</sup> se creía que esta orientación generaba menores errores de pro-

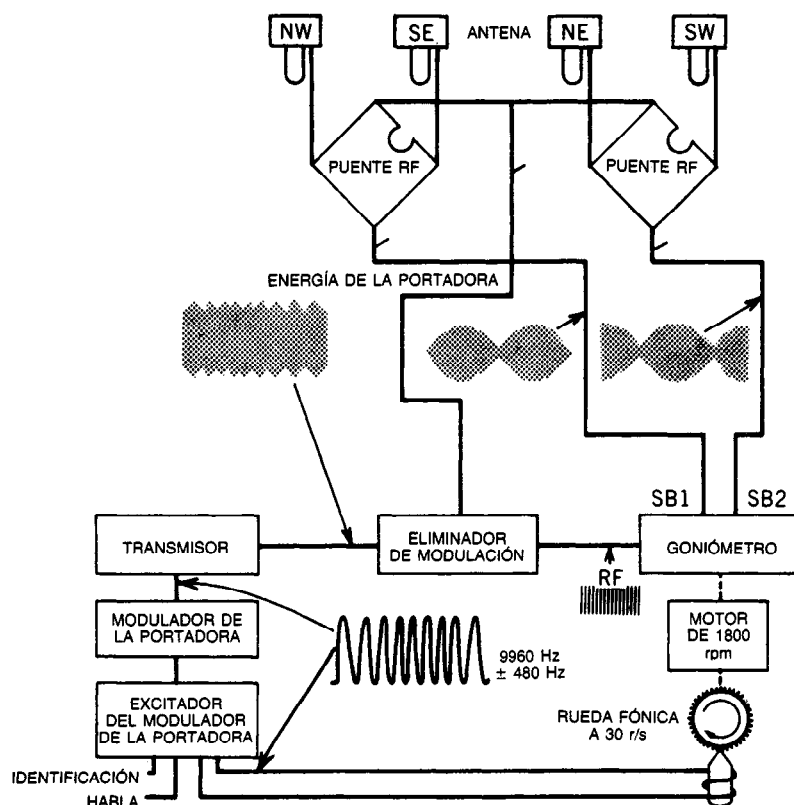
pagación de trayectoria múltiple que la polarización vertical, aunque esta conclusión no fue aceptada universalmente.

Las especificaciones de modulación son controladas por las normas de la ICAO<sup>4</sup> y de Estados Unidos.

**Diseño de la estación.** Un elemento importante en la radiación de señales VOR polarizadas horizontalmente es la antena de cuadro horizontal Alford<sup>6</sup>. Esta tiene un diagrama ("patrón") de radiación (o de intensidad de campo) similar al de un dipolo vertical, pero con polarización horizontal. Es un cuadro rectangular horizontal con lados que miden aproximadamente  $0.25 \lambda$  de longitud. Para el transmisor VOR, se disponen cuatro de estos elementos en un cuadrado de empaque compacto. Cuando dos cuadros opuestos diagonalmente reciben al mismo tiempo la señal, su diagrama de radiación es un número ocho horizontal. El diagrama del otro par es un número ocho perpendicular al primero. Cuando a cada par de cuadros se suministran dos señales moduladas en amplitud a 30 Hz, pero defasadas  $90^\circ$ , se obtiene un diagrama de radiación compuesto que es un número ocho horizontal que gira a 30 r/s. Cuando este patrón se combina en el receptor con la componente portadora de la señal de referencia que se emite desde una antena omnidireccional, el patrón se convierte en un cardioide giratorio.

En la mayor parte de las estaciones VOR existentes en la actualidad, las dos señales defasadas de 30 Hz se generan suministrando energía portadora a un radiogoniómetro capacitivo, un defasador mecánico, que se hace girar a una velocidad exacta de 1 800 rpm por medio de un motor sincrónico<sup>7</sup>. En la figura 52-3 se ilustra un diagrama de bloques de la estación VOR ordinaria, en la que se emplea modulación mecánica. La señal de  $9\,960 \pm 480$  Hz proviene de una rueda fónica impulsada por el mismo eje que el radiogoniómetro. Esta rueda tiene 332 dientes con espaciamiento cíclicamente escalonado. La señal se capta inductivamente de esta rueda, se suma a las señales de habla e identificación, y esta combinación se utiliza para modular en amplitud el transmisor. Se suministra al radiogoniómetro parte de esta señal cuya modulación ha sido eliminada por un limitador. Esta disposición preserva la relación de fase entre las dos señales. En dos puentes RF se combinan las componentes de fase de referencia y variable.





**Fig. 52-3.** VOR (radiofaro omnidireccional VHF) ordinario con antena de cuatro cuadros. Tomado de R. B. Flint, "VOR Past, Present, Future", presentado en el simposio aeroespacial nacional del Institute of Navigation, Atlantic City, NJ, reproducido con permiso.

**Diseño del receptor.** La figura 52-4 es el diagrama de bloques general de un receptor VOR. La señal compuesta es recuperada por el detector de AM. Filtros independientes extraen la voz, las señales de fase variable de 30 Hz y la subportadora de 9 960 Hz modulada en frecuencia. La señal de fase de referencia de 30 Hz es recuperada de esta subportadora por un limitador discriminador. Para conservar la exactitud del sistema es esencial un estrecho seguimiento de fase de los dos filtros pasabanda de 30 Hz.

Las antenas de las aeronaves varían desde los simples dipolos en V horizontales hasta ciertas variantes del cuadro Alford. Se prefiere un montaje elevado sobre el estabilizador vertical de la aeronave para lograr cobertura omnidireccional.

En lo que se refiere a la visualización para el piloto, aunque se dispone de indicadores de lectura de fases directa (analógica o digital) del acimut real o "radial" a la estación seleccionada, una visualización más común es el indicador de desviación del rumbo (CDI) que se ilustra en la figura 52-5. La escala de rumbo giratoria, llamada selector de rumbo (OBS), está conectada a un defasador de 30 Hz. Éste desplaza separando desde 0° (Norte) la señal de fase de referencia el número de grados que se observa en el apuntador del "rumbo seleccionado". La aguja de desviación es impulsada por

un resolutor (sincrotrigonómetro), que compara la fase de esta referencia desplazada con la señal de 30 Hz de fase variable. Por lo general una desviación de escala completa indica un error de rumbo de 10°. La aguja se mueve hacia el centro cuando la aeronave se encuentra en el rumbo seleccionado a la estación, o bien en su recíproco. Para resolver esta ambigüedad, una de las señales de 30 Hz se desplaza 90° y se compara con la otra en un segundo resolutor de alta ganancia, que activa una señal (bandera) de resolución de la ambigüedad "TO-FROM" (hacia-desde). Cuando los niveles de las señales de 30 Hz se vuelven demasiado débiles, aparece una señal (bandera) de advertencia en una segunda ventana en el instrumento.

En virtud de que se está instalando en cerca del 90% de las aeronaves de un solo motor y bimotores ligeras de Estados Unidos, el CDI simple es el indicador VOR más común; sin embargo, en aeronaves grandes se encontrará un indicador similar al indicador de situación horizontal (HSI), que se muestra en la figura 52-6. Las indicaciones se superponen sobre una carátula giratoria de brújula esclavizada al sistema de compás principal del avión. La flecha de rumbo apunta hacia el "radial" VOR seleccionado por la perilla de fijación del rumbo (290° en el ejemplo). La barra de desviación del rumbo se mueve para mostrar la po-

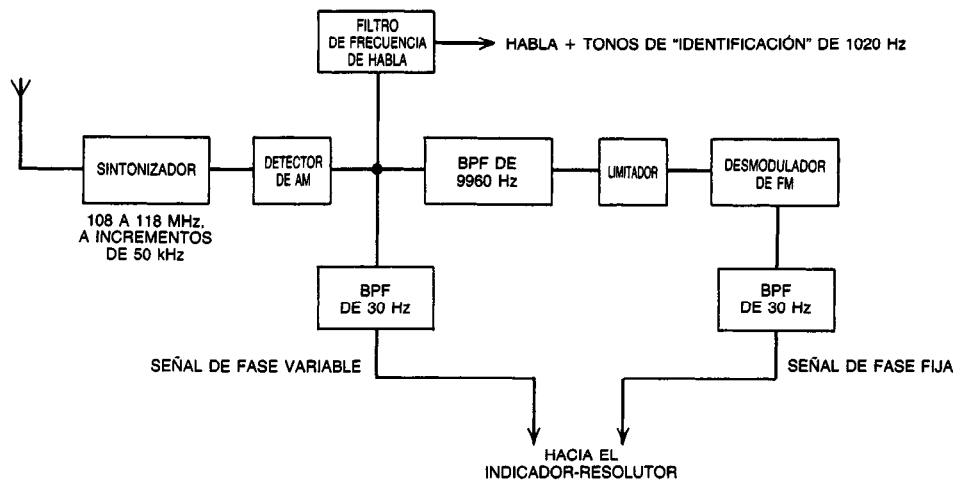


Fig. 52-4. Receptor VOR (radiofaro omnidireccional VHF). BPF = filtro pasabanda.

sición de la radial de acimut seleccionada con respecto a la aeronave simbólica fija en el centro del indicador. Cuando la barra se alinea con la flecha de rumbo, la aeronave se encuentra en el acimut seleccionado. Un par de flechas de dirección muestra si la estación VOR está en frente o detrás del avión, y corresponde a las señales "TO-FROM" del CDI. Además, aparecerán señales (banderas) de advertencia en forma prominente en el indicador cuando la intensidad de la señal sea inadecuada.

Existen muchas variaciones del HSI básico. Suelen incluirse datos de otras fuentes de navegación, por ejemplo del sistema radiogoniométrico automático (ADF), que se describe más adelante, y el sistema de aterrizaje por instrumentos (ILS) (cap. 53). Para la siguiente generación de aeronaves de transporte, la simbología será generada por computadora y se visualizará en un tubo de rayos catódicos.

El Radio Technical Committee for Aeronautics (RTCA)<sup>8</sup> ha establecido las características de rendimiento mínimo de los receptores VOR, y la Aeronautical Radio, Inc. (ARINC) ha establecido las normas del equipo para líneas aéreas. Un receptor VOR con indicador para aviones ligeros pesaría unos 2 kg, tendría un volumen de 1 750 cm<sup>3</sup> y consumiría 20 W, mientras que en el caso de una unidad para líneas aéreas estas cifras podrían ser 6 kg, 5 300 cm<sup>3</sup> y 70 W.

**Errores de sitio.** La estación VOR ordinaria requiere una ubicación cuidadosa para evitar errores angulares de acimut producidos por las reflexiones de trayectoria múltiple desde estructuras y terrenos circunvecinos. Comúnmente las antenas se colocan más o menos media longitud de onda por encima de un contrapeso (antena de contrapeso) metálico circular de unos 7 m de diámetro, que, a su vez, se eleva cerca de 5 m por encima del suelo. Algunas veces se necesitan contrapesos mucho mayores y situados a una altura mucho mayor. Para sitios más complicados se desarrolló el VOR Doppler.

**VOR Doppler.** El VOR Doppler<sup>9</sup> puede reducir parcialmente los efectos de trayectoria múltiple mediante el uso de una abertura de antena mucho mayor que en el caso de la estación ordinaria. Un arreglo circular de 50 elementos de cuadro Alford en un diámetro de 13.4 m se coloca encima de un contrapeso con diámetro nominal de 45.7 m. Si  $f_0$  es la frecuencia nominal de la estación, se conmuta secuencialmente a cada elemento de la antena una portadora a  $f_0 + 9\,960$  Hz a

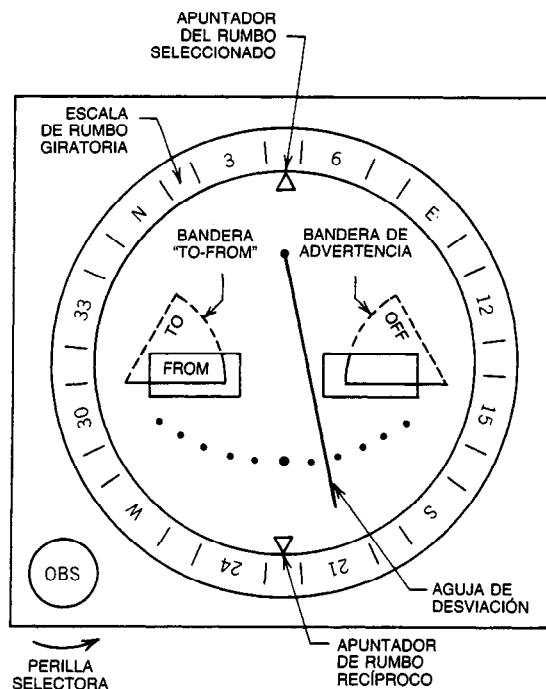


Fig. 52-5. Indicador de desviación del rumbo (CDI). OBS = selector de rumbo; TO = hacia; FROM = desde; OFF = apagado.

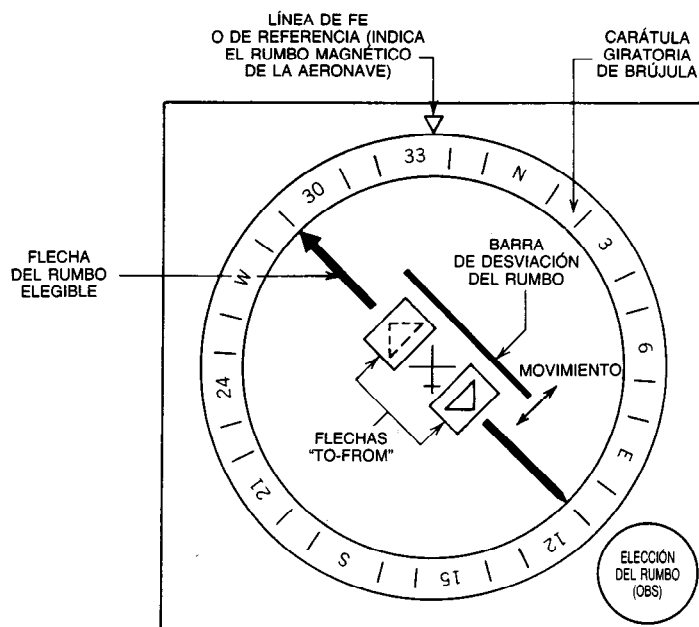


Fig. 52-6. Indicador de situación horizontal (HSI). OBS = selector de rumbo.

una velocidad de rotación de 30 r/s. Un observador situado a cierta distancia apreciará una modulación sinusoidal de frecuencia de 30 Hz aparente debida al efecto Doppler. El espaciamiento y el diámetro del anillo de la antena son tales que la desviación de frecuencia a nivel de cresta es  $\pm 480$  Hz, que corresponde a la del VOR ordinario. La fase de esta modulación de 30 Hz es función del acimut del observador desde la estación. Se genera una fase de referencia modulando en amplitud DSB con un tono fijo de 30 Hz una portadora radiada omnidireccionalmente con frecuencia  $f_0$ . Por tanto, para el VOR Doppler, las fuentes de las modulaciones de señales de 30 Hz "variable" y "de referencia" se invierten con respecto a lo que ocurre en el VOR ordinario. Sin embargo, como el receptor de la aeronave sólo mide su diferencia de fase, su desempeño es el mismo que el que se obtendría con una estación de transmisión ordinaria.

Comparado con el espectro de la señal VOR típica, el VOR Doppler genera únicamente las componentes de banda lateral superior de la subportadora de 9 960 Hz modulada en frecuencia. Algunos receptores VOR (principalmente las unidades de menor tamaño y menos costosas) han mostrado sensibilidad a este efecto, produciendo errores de fase mayores que los admisibles.

Para compensar este efecto se ha probado el VOR Doppler de doble banda lateral, en el que una portadora de  $f_0 - 9\,960$  Hz se conmuta al mismo tiempo con la de  $f_0 + 9\,960$  Hz, pero con las conexiones separadas físicamente  $180^\circ$  en torno al anillo de antenas, que inserta la banda lateral inferior. Precisamente han comenzado a instalarse en forma operacional los VOR de este tipo.

En Europa se ha implementado una técnica semejante de doble banda lateral que se conoce como VOR Doppler alternante de banda lateral<sup>10</sup>. En ésta se emplean 39 antenas de cuadro en la circunferencia. En vez de excitar al mismo tiempo cuadros diametralmente opuestos con las señales de banda lateral superior e inferior, se les excita en forma alternativa. La conformación de la potencia se utiliza de modo que cuando la potencia de la señal sea creciente en un elemento de la antena, sea decreciente en el cuadro adecuado en el lado opuesto de la circunferencia.

**Parámetros y exactitudes del sistema.** En el espacio aéreo de Estados Unidos, las estaciones VOR están dispuestas en tres clases, cada una con su volumen de servicio estándar (SSV)<sup>5</sup>:

Designador de clase	Límites de altitud y distancia de SSV
T (terminal)	Hasta 12 000 pies sobre el nivel de tierra de la estación; radio de 25 nmi
L (baja altitud)	Hasta 18 000 pies; radio de 40 nmi
H (elevada altitud)	Hasta 14 500 pies; radio de 40 nmi 14 500 a 18 000 pies; radio de 100 nmi 18 000 a 45 000 pies; radio de 130 nmi 45 000 a 60 000 pies; radio de 100 nmi

También se pueden definir volúmenes de servicio extendidos (ESV) para estaciones VOR. El volumen de servicio operacional (OSV) es la suma de SSV y ESV de esa estación. Dentro del OSV la mínima densidad de potencia de señal será  $-120 \text{ dBW/m}^2$  (disponibilidad el 95% del tiempo).

Para cumplir estos requisitos, la potencia de los transmisores VOR está comúnmente en el intervalo de 10 a 200 W.

Los métodos que emplean los diversos Estados miembros de la ICAO para definir la exactitud varían, pero en Estados Unidos el error del sistema agregado para el VOR se da como  $\pm 4.5^\circ$ , que es el valor cuadrático resultante de las siguientes componentes de error (valores de probabilidad del 95%):

1. **Error de señal radial:**  $\pm 1.4^\circ$ . Un límite de tolerancia monitorizado por la Federal Aviation Agency verifica el vuelo dentro de un volumen de servicio operacional de la estación VOR.
2. **Error de componentes de a bordo:**  $\pm 3.0^\circ$ . Error debido al sistema receptor (exclusivo de los indicadores). Este valor es establecido por receptores de aviación generales de bajo costo. El equipo moderno de las líneas aéreas reducirá esto en un factor de cuatro.
3. **Error de instrumentación:**  $\pm 2.0^\circ$ . Errores debidos al indicador (incluido el error de ajuste) y el acoplador de piloto automático, cuando se utiliza. Es significativamente menor en el caso de pantallas digitales.
4. **Errores técnicos de vuelo:**  $\pm 2.3^\circ$ . Error debido a la imprecisión del pilotaje. Se reduce con el acoplador de piloto automático.

#### Equipo radiotelemétrico (DME)

El DME estándar de FAA e ICAO es un sistema de telemetría por pulsos. Cada aeronave interroga a un transpondedor (radiofaro de respuesta) de tierra y mide el retardo de viaje de ida y vuelta entre los pulsos de interrogación y de respuesta. Se creó originalmente como el elemento de medición del sistema TACAN (navegación aérea táctica); no obstante, las funciones de DME pueden implementarse en forma totalmente independiente<sup>11</sup>.

Las interrogaciones de las aeronaves se efectúan en una de 126 frecuencias espaciadas 1 MHz, de 1 025 a 1 150 MHz. La respuesta del radiofaro de tierra está en una frecuencia desviada 63 MHz de la interrogación. Cuando la respuesta es 63 MHz inferior, la disposición recibe el nombre de canal X, y cuando es mayor, canal Y. Por tanto, existen 252 frecuencias de respuesta que van de 962 MHz a 1 213 MHz. Las señales de aire y de tierra se radian con polarización vertical, y los diagramas de directividad de la antena de tierra se hacen lo más omnidireccionales que sea posible. Cada interrogación y respuesta consta de un par de pulsos. El espaciamiento de pulsos es como se indica en la tabla del comienzo de la columna siguiente:

La tolerancia en el espaciamiento es de 0.25  $\mu\text{s}$ . Cada pulso tiene distribución gaussiana a fin de minimizar la ocupancia espectral, con ancho de media

Canal	Espaciamiento de pulsos interrogadores del avión	Espaciamiento de la respuesta de tierra
X	12 $\mu\text{s}$	12 $\mu\text{s}$
Y	36 $\mu\text{s}$	30 $\mu\text{s}$

amplitud de  $3.5 \pm 0.5 \mu\text{s}$  (ref. 4, vol. 1, sec. 3.5). El radiofaro de respuesta (transpondedor) de tierra está diseñado para introducir un retardo fijo de 50  $\mu\text{s}$  entre la decodificación correcta de un par de pulsos y la retransmisión de otro par de pulsos. Este retardo fijo es restado por el equipo de la aeronave en su estimación del tiempo de viaje de ida y vuelta.

En la actualidad sólo los canales X se utilizan realmente en las aeronaves civiles, aunque el uso de canales Y ha sido autorizado por el Department of Transportation de Estados Unidos y por la ICAO, y la mayor parte de las instalaciones DME de aeronaves civiles ya cuentan con este recurso.

**Adquisición y seguimiento.** Cada transpondedor o radiofaro de respuesta DME de tierra operará, desde luego, en forma simultánea con muchas aeronaves dentro de su área de cobertura. Como cada aeronave que interroga a ese radiofaro utiliza la misma frecuencia y formato de señal, y las interrogaciones de las aeronaves no están sincronizadas de ninguna manera, cada aeronave tiene el problema de identificar las respuestas a sus interrogaciones de aquellas respuestas a las interrogaciones de otras naves. Para lograr este fin, la velocidad de interrogación de cada aeronave se distorsiona deliberadamente en forma aleatoria. Después, el receptor DME busca un patrón de pares de pulsos devueltos que ocurren con un retardo fijo a partir del tiempo de interrogación de la aeronave. Cuando la compuerta de búsqueda de la nave se coloca correctamente respecto al tiempo de interrogación, se detecta una réplica en respuesta a la mayor parte de las interrogaciones "de la aeronave propia". Las respuestas de las interrogaciones de otras aeronaves variarán al azar con respecto a este tiempo de conteo o "de compuerta".

Para la búsqueda inicial, la aeronave hace un promedio de aproximadamente 150 interrogaciones por segundo. La compuerta de intervalo se desplaza hacia afuera a partir del retardo de intervalo cero hasta que se detecta una elevada proporción de pares de respuestas. Esto suspende la búsqueda y da inicio al modo de seguimiento. En éste, la velocidad promedio de interrogación se reduce a un valor menor que unas 20 interrogaciones por segundo. Ahora la compuerta se re-coloca en forma continua, con el uso de un ciclo de seguimiento predictivo, a fin de mantener los pulsos devueltos en el centro de la compuerta a medida que cambia la distancia del avión. Los ciclos de predicción hacen posible el seguimiento aun cuando falte una elevada proporción de los pulsos de respuesta. Con el equipo aéreo moderno se pueden practicar velocidades de interrogación tan bajas como tres por segundo.

El DME mide la distancia real entre la aeronave y el radiofaro, en vez de la distancia a tierra. El efecto práctico de esta diferencia es insignificante, salvo cuando la distancia entre el avión y la estación de tierra es de unas cuantas millas náuticas.

**Implementación del radiofaro de tierra.** El radiofaro de respuesta de tierra recibe y transmite a un par de frecuencias fijas separadas por 63 MHz. La antena del radiofaro suele constar de dipolos apilados verticalmente, lo que produce un patrón omnidireccional en el plano horizontal. La ganancia es cuando menos de 4 dB en un solo dipolo; algunas veces se emplean valores hasta de 9 dB. Después de cada interrogación con éxito el receptor de tierra se inhibe durante un retardo aproximado de 60  $\mu$ s. Esto tiene el fin de evitar que los ecos demorados de trayectoria múltiple activen la interrogación directa. En terreno montañoso muy accidentado, este retardo de 60  $\mu$ s quizá tenga que incrementarse hasta 150  $\mu$ s.

Los radiofaros están diseñados para operar a una velocidad de respuesta de 2 700 pares de pulsos por segundo. Esto corresponde al manejo de 100 aeronaves, el 5% de ellas en el modo de búsqueda (150 interrogaciones por segundo) y el 95% en el modo de seguimiento (un promedio de 20 interrogaciones por segundo). Esta razón de pares de pulsos se mantiene aunque haya presentes menos interrogaciones o ninguna, al incrementar en forma automática la ganancia del receptor de tierra hasta que el radiofaro active su propio ruido de entrada. La ganancia se ajusta dinámicamente de manera que el número total de puntos transmitidos debido a las interrogaciones y al ruido aleatorio promedie 2 700 pares de pulsos por segundo. Esta disposición hace posible que:

1. El transmisor opere en un régimen de trabajo constante.
2. El control automático de ganancia del receptor de la aeronave tenga un número constante de pulsos que integrar, aun antes de comenzar a buscar interrogaciones.
3. La degradación sea justa en caso de saturación por un número elevado de aeronaves: primero se ignoran las interrogaciones de la aeronave más distante.

El intervalo de variación de ganancia del radiofaro es de 50 dB o mayor.

**Identificación.** Las especificaciones de la FAA y la ICAO requieren radiofaros DME para multicanalizar códigos de identificación audibles junto con las señales de navegación. En intervalos de más o menos 30 s se suprime la transmisión de pares de pulsos de respuesta y aleatorias, mientras que se transmiten dos a tres caracteres de identificación en clave Morse. Los símbolos duran 0.125 s en el caso del punto y 0.375 s en el de la raya, y se envían como pares de pulsos a una velocidad constante de 1 350 pares de pulsos por segundo. Esta señal se detecta en la aeronave por un circuito sintonizado, y se alimenta el sistema de distribución de

audio como un tono en clave Morse de 1 350 Hz. En los espacios entre los puntos y rayas se intercalan respuestas regulares y pulsos aleatorios. La transmisión de cada grupo de códigos dura menos de 5 s.

**Implementación en la aeronave.** La figura 52-7 es un diagrama de bloques del DME aéreo típico. Las frecuencias de transmisión y recepción siempre están separadas 63 MHz, de manera que el uso de una frecuencia intermedia de 63 MHz hace posible que una fuente de frecuencia actúe como excitación del transmisor y como oscilador local del receptor. Las antenas de las aeronaves son segmentos verticales de un cuarto de onda, que para las frecuencias de los DME tienen unos 7 cm de longitud. Las potencias de cresta de pulso del transmisor varían desde 100 W para las unidades más simples hasta 1 kW o más para el equipo de líneas aéreas.

En todos los equipos modernos, el retardo del viaje de ida y vuelta se mide contando digitalmente una señal de reloj precisa. Después de la sustracción de tiempo de retardo fijo de 50  $\mu$ s del radiofaro, cada milla náutica corresponde a unos 12  $\mu$ s de retardo del viaje de ida y vuelta. Aunque el máximo alcance utilizable es determinado por las limitaciones de la trayectoria óptica (rectilínea), el alcance máximo de medición del equipo suele ser de 200 millas náuticas, si bien algunas unidades diseñadas para aeronaves pueden medir hasta 400 millas náuticas.

Las pantallas digitales presentan una lectura de la distancia verdadera hasta el radiofaro seleccionado en millas náuticas y décimas. Una segunda pantalla digital se obtiene de los ciclos de seguimiento de la velocidad e indica la velocidad de cambio de la distancia en nudos. Cuando la aeronave sigue el rumbo directamente hacia o desde el transpondedor o radiofaro de respuesta (el caso usual de las aeronaves civiles), ésta es una aproximación a la velocidad absoluta (respecto a tierra). Suele presentarse una tercera pantalla digital, que expresa el "tiempo de arribo hasta la estación" directamente en minutos, siempre que la velocidad de cambio de la distancia sea negativa (es decir, durante el vuelo hacia el radiofaro).

La RTCA establece unas normas de rendimiento mínimo para el DME de aeronaves<sup>13</sup>. Un DME básico de una aeronave ligera pesaría 1.5 kg, ocuparía 1 650 cm<sup>3</sup> y consumiría cerca de 30 W. En el caso de unidades de líneas aéreas estas cifras se aproximarían más a 7.5 kg, 11 000 cm<sup>3</sup> y 55 W.

**Parámetros y exactitudes del sistema.** Los volúmenes de servicio operacionales (estándar más extendido) para radiofaros DME (ref. 5, cap. 4) se definen exactamente en la misma forma que para las estaciones VOR. Los volúmenes utilizables se definen como aquellos para los cuales la densidad de potencia en la antena de la aeronave es la siguiente:

- Más de 18 000 pies sobre el nivel de tierra, -91.5 dBW/m<sup>2</sup>
- Menos de 18 000 pies sobre el nivel de tierra, -86.0 dBW/m<sup>2</sup>

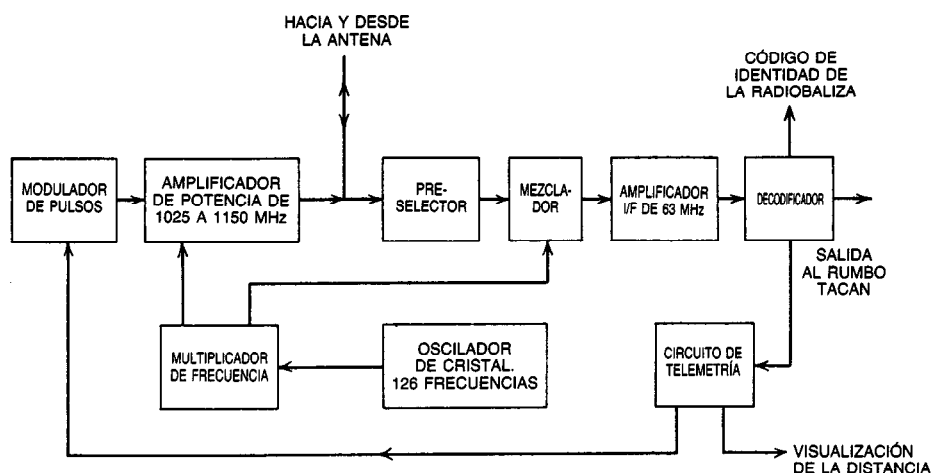


Fig. 52-7. Diagrama de bloques de la radiotelegrafía de avión (DME). Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.

Éstos son promedios de las crestas de los pulsos, y son cifras de disponibilidad el 95% del tiempo. Para obtener estos valores, las potencias de cresta del radiofaro DME de tierra varían de 10 kW en el caso del servicio en ruta de larga distancia a 1 kW en el de radiofaros diseñados para servicio sólo de terminal (radio de 30 nmi o menos) o para radiofaros instalados en asociación con un sistema de aterrizaje por instrumentos (ILS) (cap. 53).

A máxima ganancia, la especificación de sensibilidad del receptor del radiofaro requiere una eficiencia de respuesta del 70% cuando la densidad de potencia de cresta incidente de una interrogación de la aeronave es  $-101.5 \text{ dBW/m}^2$ .

Las exactitudes que se requieren del DME están dadas en las especificaciones de la ICAO y de la FAA. La asignación de error debido a todas las funciones del radiofaro de tierra es un máximo de 0.1 nmi (error de retardo aproximado de  $1 \mu\text{s}$ ). La combinación (valor cuadrático resultante) de este error con el que se puede atribuir al equipo del avión es el error total de distancia, que se presenta al piloto y no debe exceder de 0.5 nmi o el 3% de la distancia total, lo que sea mayor. En la práctica el error real es, con el 95% de confianza,  $\pm 0.2 \text{ nmi}$  o bien el 0.25% de la distancia real, lo que sea mayor. Como parte del sistema de aterrizaje por microondas (MLS) se está desarrollando un DME compatible, pero que tiene una exactitud significativamente mayor en cortas distancias (cap. 53).

#### Navegación aérea táctica (TACAN)

La navegación aérea táctica (TACAN) es un sistema de navegación con fuente puntual (rho-theta) utilizado por los militares de Estados Unidos, la OTAN y otras fuerzas armadas del mundo occidental. Su desarrollo data de 1948 y su producción, de 1953<sup>14</sup>.

Un radiofaro de tierra (o a bordo de un barco) TACAN consta de un radiofaro DME, ya descrito, con modulación de amplitud agregada a su respuesta y pul-

sos transmitidos en forma aleatoria a fin de proporcionar información de acimut. Los formatos y frecuencias de los pulsos radiotelemétricos son los mismos que se definen en las especificaciones de la FAA de la ICAO para el DME (es decir, 126 canales X y 126 canales Y). En el documento de normas militares TACAN de Estados Unidos se dan especificaciones de las funciones de medición del acimut<sup>15</sup>.

**Formato de señales del acimut.** Los pulsos que se transmiten son modulados en amplitud por una combinación de dos ondas de forma sincronizadas generadas por un sistema de antena que gira a 15 r/s. La primera componente es una modulación de amplitud (AM) sinusoidal a 15 Hz (una vez por revolución), y la segunda es una AM sinusoidal a 135 Hz (nueve veces por revolución). La rotación es en el sentido del reloj, vista desde arriba del radiofaro. Cuando el radiofaro está en un ángulo de acimut magnético y desde un observador, la envolvente de los pulsos detectados vista por ese observador está dada por:

$$y(t) = 1 + A \sin(2\pi ft - \gamma) + B \sin(18\pi ft - 9\gamma)$$

donde  $y(t)$  = envolvente detectada compuesta normalizada

$A$  = modulación de la componente de 15 Hz =  $0.21 \pm 0.09$

$B$  = modulación de la componente de 135 Hz =  $0.21 \pm 0.09$  (la suma de  $A$  y  $B$  no debe exceder de 0.55)

$f$  = frecuencia de rotación del patrón (diagrama) =  $15 \text{ Hz} \pm 0.23\%$

$t$  = tiempo transcurrido desde algún "tiempo de referencia" definido

La posición acimutal se determina midiendo, en  $y(t)$ , la fase de las señales de 15 y 135 Hz. A diferencia del VOR, en el cual  $1^\circ$  de fase eléctrica corresponde a  $1^\circ$  de acimut, para el TACAN  $9^\circ$  de fase eléctrica de la

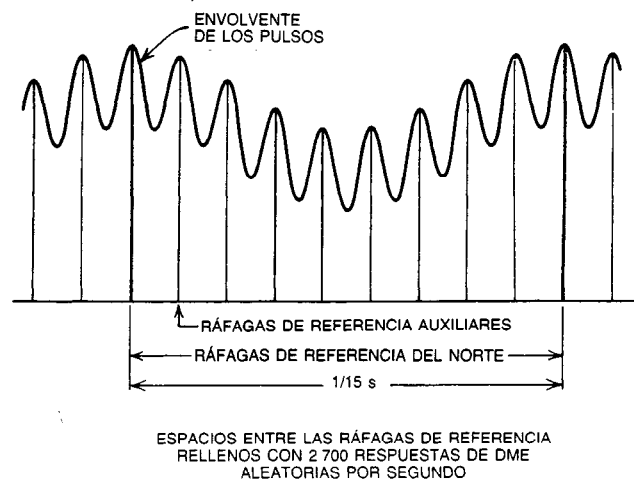


Fig. 52-8. Señal de navegación aérea táctica (TACAN) (al este de la estación).

señal de 135 Hz corresponden a 1° de acimut. Como la señal de 135 Hz tiene nueve ambigüedades por revolución (en intervalos de 40°), es necesaria la fase de la componente de 15 Hz para resolverlas.

Dos conjuntos de señales de identificación definen el "tiempo de referencia" a partir del cual se mide  $t$  en la ecuación dada. El primer conjunto se denomina grupo de referencia principal y se transmite cuando la cresta de la modulación de 15 Hz apunta directamente al este de la estación (ésta se conoce como la señal de referencia "norte"). El segundo conjunto de señales se

transmite cada vez que la cresta de la modulación de 135 Hz apunta al este y se denomina grupo de referencia auxiliar. Por tanto, existen nueve grupos auxiliares por cada grupo principal, salvo que, como las crestas de las dos modulaciones están sincronizadas cuando las crestas de modulación de 15 y 135 Hz apuntan al este, las señales de referencia auxiliares se suprimen en favor de la señal de referencia principal. La figura 52-8 ilustra la onda de forma envolvente de los pulsos que sería vista por un usuario dirigido magnéticamente al este de la estación TACAN.

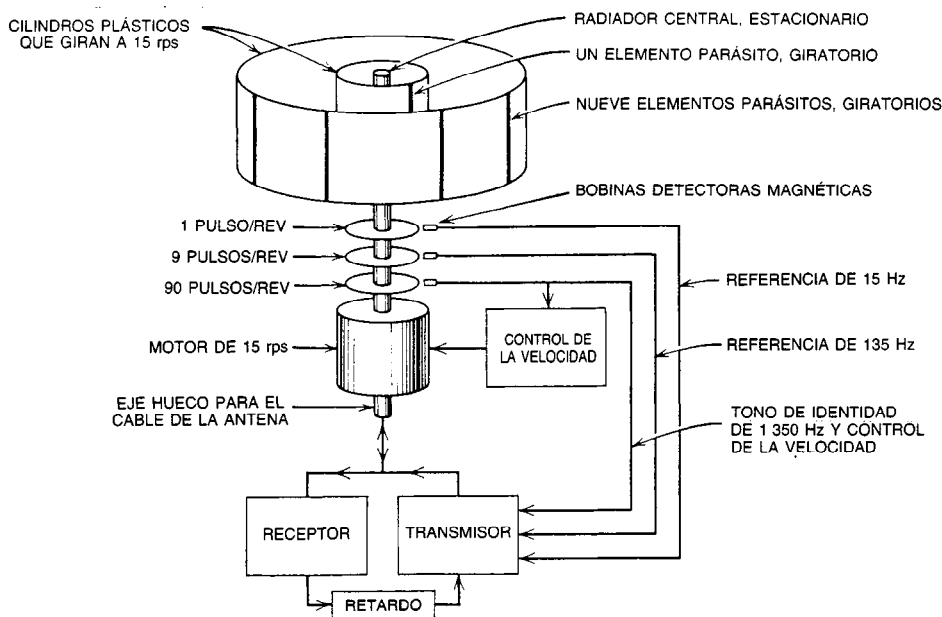
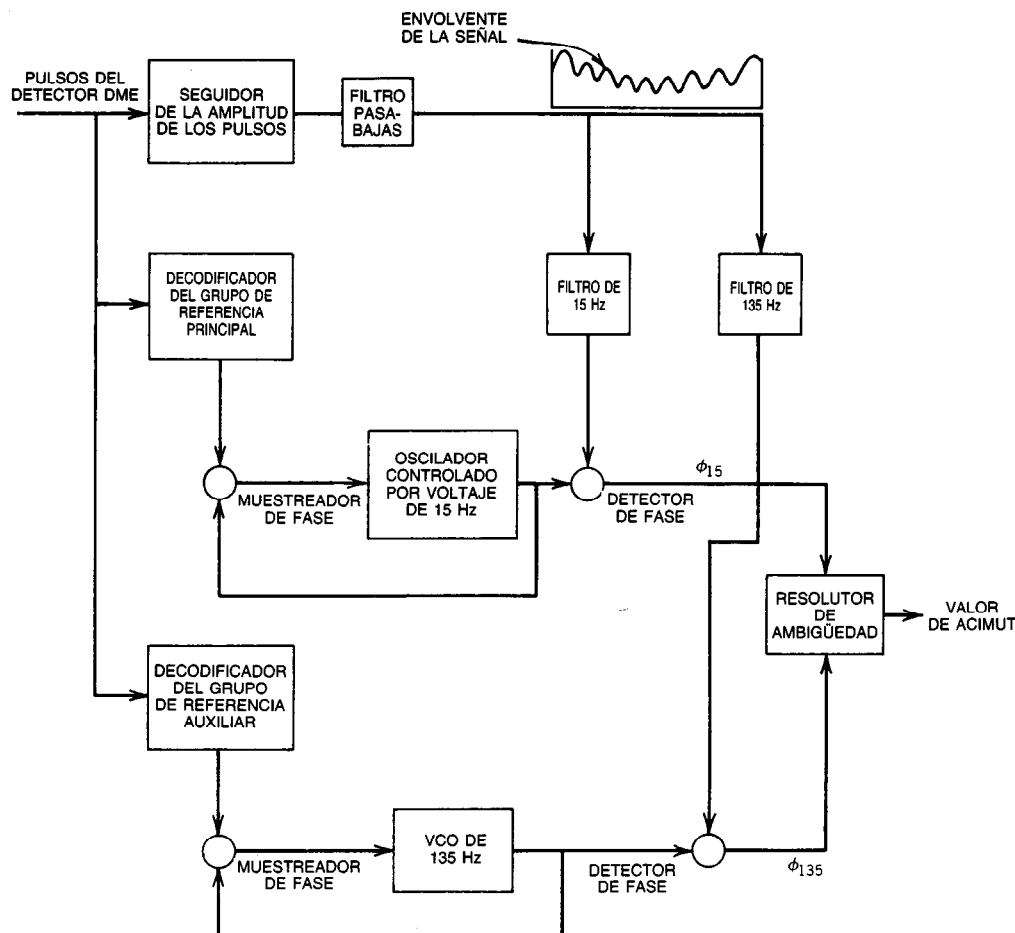


Fig. 52-9. Radiofaro de tierra de navegación aérea táctica (TACAN). Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.



**Fig. 52-10.** Medición de acimut de navegación aérea táctica (TACAN) a bordo. DME = equipo radiotelemétrico; VCO = oscilador controlado por voltaje.

Los grupos de referencia principal y auxiliares son generados por grupos de pulsos con espaciado de pulsos específico, como se define en la referencia 15. Durante la transmisión de los pulsos de los grupos de referencia principal o auxiliares, se suprimen los pulsos aleatorios, de identidad y de respuesta de distancia.

**Configuración del radiofaro.** Un radiofaro de tierra TACAN es un radiofaro DME más un sistema de antena de modulación y circuitos para generar los grupos de pulsos de referencia. La figura 52-9 ilustra la configuración. La mayor parte de los radiofaros TACAN en uso tienen sistemas de antena mecánicamente rotatorios. Dos conjuntos de elementos parásitos se hacen girar alrededor del radiador central a 15 r/s para producir la modulación de amplitud compuesta. Los cilindros plásticos interior y exterior que transportan estos conjuntos de elementos miden unos 15.25 cm y 91 cm de diámetro, respectivamente. Los pulsos activadores de los grupos de pulsos de referencia principal, auxiliares y de identidad de 1 350 Hz se obtienen de

tres discos que giran sobre el mismo eje. Para el uso a bordo de barcos, la antena del TACAN se estabiliza mecánica o electrónicamente para compensar el aporamiento y balanceo del barco, de manera que las señales en el espacio definan con exactitud el rumbo magnético referido al barco.

**Antenas de estado sólido.** Se hallan en servicio antenas de estado sólido moduladas electrónicamente. En una versión (Model AT-100, Rantec, división de Emerson Electric) se emplean 36 dispositivos de antenas múltiples verticales dispuestos en un círculo de 1.5 m de diámetro. Cada dispositivo mide 3.15 m de alto y consta de 12 radiadores de cavidad. Una red de alimentación acimutal y generadora de modulación produce corrientes de excitación en la fase correcta para generar el diagrama giratorio de nueve lóbulos. Este sistema pesa 350 kg, en contraste con los más de 1 000 kg de la antena rotatoria ordinaria. Y, además, sólo requiere unos 100 W de potencia de activación, en lugar de los 5 kW de las antenas de rotación mecánica.



**Equipo de a bordo.** En la figura 52-10 se muestra la configuración funcional del equipo para decodificar señales de acimut del TACAN. Cuando se decodifican los grupos de pulsos de referencia principal y auxiliares, se generan pulsos de muestra que se emplean para sincronizar osciladores de 15 y 135 Hz. La fase de las señales de 15 y 135 Hz detectadas se compara por separado con estas referencias. La ambigüedad de nueve veces por revolución en la señal de 135 Hz se resuelve por medio de un proceso similar al que se desarrolla a continuación:

Sean  $\phi_{15}$  y  $\phi_{135}$  las dos fases medidas (definidas sobre  $\pm 180^\circ$ ). Mídase el entero  $K$  como:

$$\left[ \frac{\phi_{15} + 20^\circ \times \text{sign}(\phi_{15})}{40^\circ} \right]$$

donde  $[]$  representa la parte entera. Por tanto, el acimut medido es:

$$\frac{\phi_{135}}{9} + K \times 40^\circ$$

Sólo ocurrirán errores de ambigüedad si el valor medido de  $\phi_{15}$  tiene un error de más de  $\pm 20^\circ$ .

**Pantallas de a bordo.** La visualización de acimut del TACAN es un apuntador rotatorio activado por la señal de la estación TACAN seleccionada sobre  $0 \rightarrow 360^\circ$ . Cuando el indicador está esclavizado a la brújula giroscópica (giroscopio direccional o girocompás) de la aeronave, la carátula y el apuntador giran y muestran:

1. El rumbo de la aeronave.
2. El acimut del TACAN.

Este instrumento se conoce como indicador radio-magnético o indicador combinado (RMI). Las aeronaves militares modernas, que utilizan computadoras de navegación, emplean datos adquiridos al mismo tiempo de varias estaciones TACAN (dentro del alcance), los combinan con los de otras fuentes (p. ej., inerciales) y producen una visualización geográfica, que podría ser como coordenadas de latitud y longitud, rumbo y distancia a un punto cualquiera seleccionado, o bien un mapa gráfico en movimiento.

**Parámetros y exactitudes de las señales.** Debido a la afinidad de las señales de acimut del TACAN con las del DME, los volúmenes de servicio y densidades de potencia requeridos son tal como se especifican para el servicio exclusivo de DME. La norma nacional de Estados Unidos para el TACAN especifica  $\pm 3^\circ$  (95 % de probabilidad) como el error admisible, incluyendo errores de ubicación, decodificación y exhibición<sup>5</sup>. En la práctica, la exactitud del acimut es significativamente mayor. Se han demostrado errores totales del sistema de menos de  $\pm 0.5^\circ$  mediante el uso de estaciones de tierra bien ubicadas. La tolerancia de rumbo de la estación de tierra es  $\pm 1^\circ$  para la componente de 135 Hz y  $\pm 4.5^\circ$  para la componente de 15 Hz.

**TACAN entre estaciones aéreas.** Existe un modo especial de TACAN para la localización entre aeronaves militares cooperativas en el aire. En la referencia 15 se indican sus características. Las frecuencias de transmisión y recepción siguen estando separadas por 63 MHz, pero están dispuestas en forma diferente de las que se utilizan en la operación entre aire y tierra. La respuesta entre estaciones aéreas es un solo pulso y no un par de ellos. En este modo no se dispone de datos de acimut intrínsecos, salvo que se pueden emplear las antenas direccionales de la aeronave para medir ángulos de llegada de los pulsos de respuesta.

#### VOR/DME y VORTAC

Los radiofaros de tierra DME para vuelos civiles están colocados con estaciones VOR (VOR/DME). Cuando el DME es asimismo parte de un radiofaro TACAN militar, la disposición se conoce como VORTAC. La figura 52-11 ilustra la configuración usual. La antena TACAN está montada encima del contrapeso del VOR, en el centro con respecto a las cuatro antenas VOR de cuadro Alford. Un radomo o cúpula de antena de forma cónica cubre el equipo. Esta disposición minimiza la distorsión de los diagramas de radiación del VOR por la antena del TACAN. Si el VOR es del tipo Doppler, la disposición coaxial produce una excesiva distorsión de las señales del VOR. En este caso las antenas del VOR y TACAN están separadas hasta 80 m. En la práctica esta separación entre los orígenes del VOR y DME es insignificante.

En la figura 52-11 se ilustra el uso de una estación VORTAC por aeronaves civiles y militares para obtener información de distancia y rumbo. La modulación de amplitud TACAN de los pulsos DME es ignorada por los receptores DME de las aeronaves civiles. La colocación permite que los dos conjuntos de usuarios, con diferentes sistemas de navegación, tengan cabida en una estructura de ruta de control del tráfico aéreo común.

**Asignación de canales.** Toda frecuencia VOR se "parea" con un número de canal DME o TACAN. Cuando el piloto selecciona una frecuencia VOR específica, el DME suele sintonizarse automáticamente a la frecuencia adecuada asociada con este VOR. Estos pareamientos de frecuencia han sido establecidos por convenio internacional y promulgados en la norma correspondiente de Estados Unidos<sup>5</sup>. Existen más canales TACAN/DME que frecuencias VOR asignadas. Algunos canales TACAN/DME seorean con canales de VHF utilizados por el sistema de aterrizaje por instrumentos (ILS) (cap. 53).

El uso de dos auxiliares de radionavegación diferentes para la medición del rumbo, con cobertura y exactitud similares, podría parecer una duplicación innecesaria. La longitud de onda más corta de TACAN hace posible que el radiofaro de tierra sea mucho menor, y más fácilmente recolocable, que el radiofaro VOR con su requisito de preparación y nivelación extensas del sitio de ubicación. Los radiofaros VOR tampoco son adecuados para el uso en barcos. Por tanto, el TACAN cumple los requisitos especiales de las fuerzas militares.

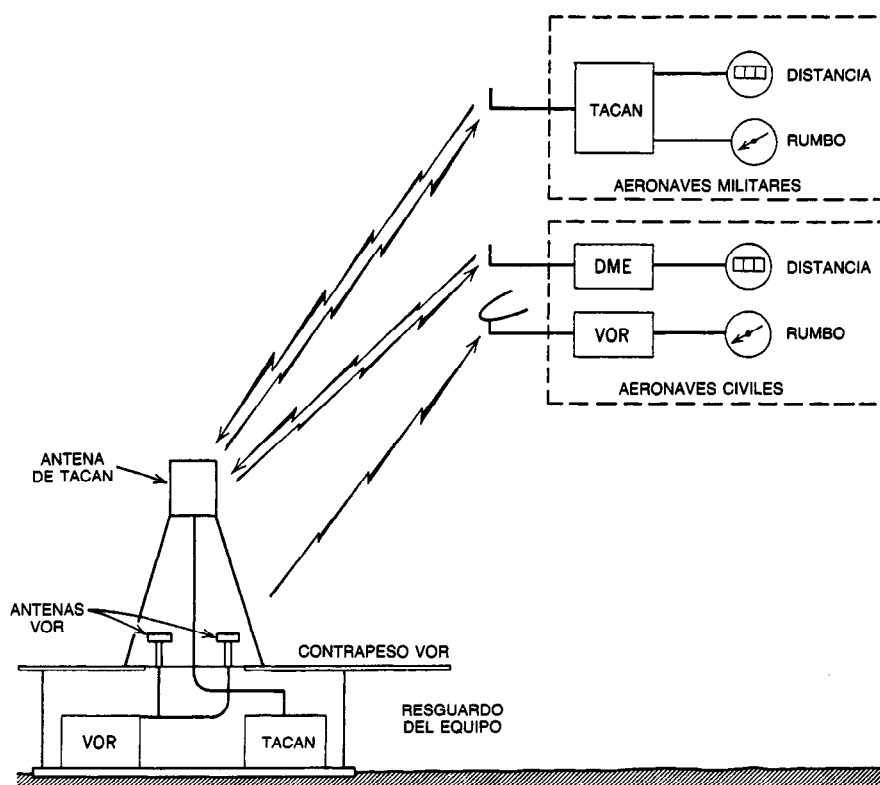


Fig. 52-11. Configuración VORTAC. TACAN = navegación aérea táctica; VOR = radiofaro omnidireccional VHF; DME = equipo radiotelemétrico. Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.

El receptor VOR de la aeronave, por otra parte, puede ser un instrumento mucho más simple y económico que el receptor TACAN más básico (en particular en el momento del despliegue inicial de los dos sistemas). Además de ofrecer datos de navegación exactos a la aeronave de transporte, el sistema VOR/DME hace posible que grandes cantidades de aeronaves de menor tamaño utilicen las vías aéreas y los sistemas de control del tráfico aéreo a un costo razonable.

#### Radiofaros no direccionales (NDB): radiogoniómetros automáticos (ADF)

Un radiofaro no direccional (NDB) es un transmisor de tierra que emite una señal esencialmente desmodulada por medio de una antena omnidireccional. La aeronave utiliza un sistema de antena direccional para rastrear esta señal y producir una indicación del rumbo del NDB respecto a la nariz de la nave. La versión moderna de este equipo de avión se conoce como radiogoniómetro automático (ADF). A diferencia del VOR o TACAN, el sistema NDB/ADF no da una indicación directa del acimut del avión respecto al radiofaro; ésta se obtiene al combinar la lectura ADF con la de la brújula magnética de la aeronave. Tampoco se dispone de información de distancia.

En Estados Unidos, los NDB rara vez se utilizan como el sistema auxiliar principal de la navegación

en ruta, salvo en Alaska. En algunas partes del mundo se pueden utilizar en combinación con el VOR. En los países en desarrollo, los NDB son probablemente los únicos sistemas auxiliares en tierra disponibles para la navegación de aviones. Los NDB se utilizan asimismo como complemento en el sistema de aterrizaje por instrumentos (ILS). Cuando se emplean de esta forma se denominan radiofaros de localización por brújula.

En Estados Unidos, los NDB tienen frecuencias asignadas en la banda de 200 a 415 kHz. La norma de la ICAO permite el uso de frecuencias en el intervalo de 200 a 1 750 kHz<sup>4</sup>. La mayor parte del equipo ADF de avión cubre este intervalo. Esto hace posible asimismo el uso para la navegación de los radiofaros de la marina (banda de 285 a 325 kHz) y también de las estaciones de radiodifusión comerciales en la banda AM estándar. Cada NDB debe transmitir una señal de identificación de dos o tres letras en clave Morse, que se repite cuando menos cada 30 segundos. Esto suele lograrse modulando en amplitud la portadora con un tono de audio manipulado de 1 020 Hz (en ocasiones 400 Hz). Fuera de Estados Unidos, algunos NDB de baja potencia transmiten los códigos de identificación manipulando por encendido y apagado la portadora no modulada. Muchos transmisores NDB también se emplean para la transmisión oral simultánea de información meteorológica para la aviación.

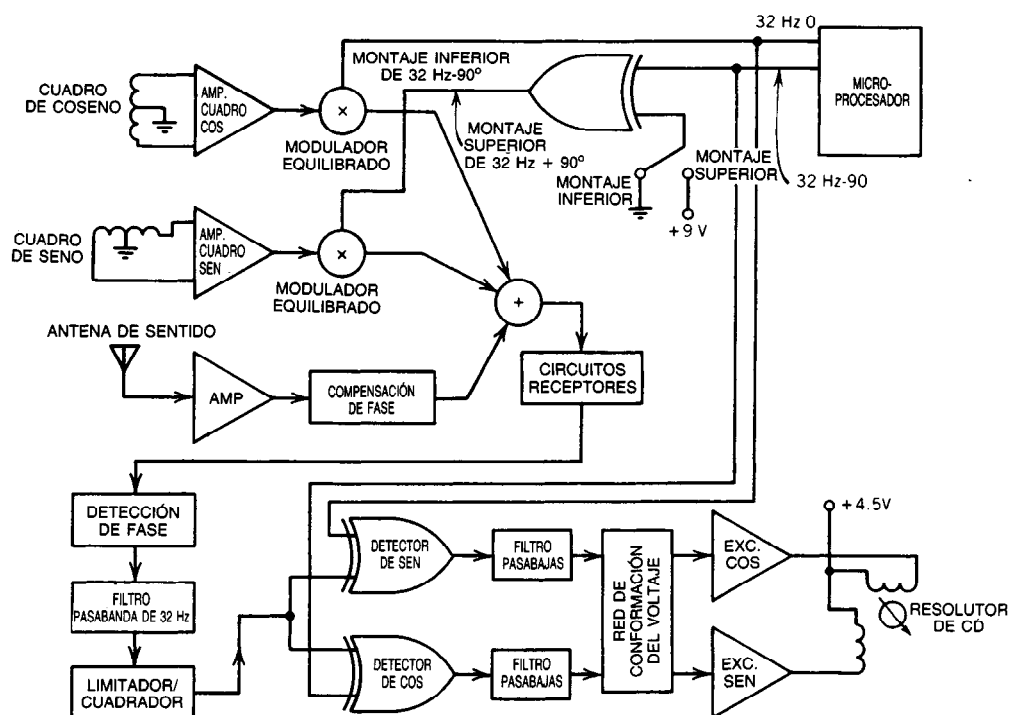


Fig. 52-12. Sistema apuntador KR87 del radiogoniómetro automático (ADF). EXC. = excitador. De King Radio, Olathe, Kansas, con permiso.

**Equipo de tierra.** El transmisor de tierra emite la señal mediante una antena omnidireccional polarizada verticalmente. La estabilidad de frecuencia está especificada por la ICAO como de 0.01% (0.005% para algunas estaciones de alta potencia). Este valor se sobrepasa sin dificultad con el equipo moderno. La potencia del transmisor varía de 20 W a 5 kW.

**Equipo de a bordo.** La operación del equipo radiogoniómetro se basa en dos antenas acopladas en fase, la antena de "cuadro" y la antena de "sentido". La primera era originalmente sólo un cuadro de varias vueltas de alambre que podía hacerse girar alrededor de un eje vertical y orientarse para una "anulación" en la señal seleccionada. Un apuntador, esclavizado a la rotación del cuadro, indicaba la dirección de la señal. Debido al diagrama horizontal en forma de ocho de la antena de cuadro se introdujo la antena de sentido, que tiene un patrón (diagrama) omnidireccional, a fin de resolver las ambigüedades de anulación de los 180°. El patrón resultante de antenas de cuadro y de sentido combinadas y puestas en fase correctamente es un cardioide, que sólo tiene una anulación.

En el equipo moderno, la antena de cuadro consta de un par de antenas devanadas fijas de placas de ferrita montadas a lo largo de los ejes transversal y longitudinal de la aeronave. Éstas reciben el nombre de antenas de cuadro seno y de cuadro coseno, debido a que los voltajes inducidos en ellas son proporcionales al seno y al coseno del ángulo de llegada de la señal.

Una vez más, un ángulo de fase arbitrario se debe resolver mediante el uso de una antena de sentido. Ésta puede ser un alambre largo que va desde el fuselaje a la parte superior del plano de deriva o bien, en aeronaves de alta velocidad, una antena de placa de campo E, por lo general montada en la parte inferior del fuselaje en una configuración al ras o casi al ras. De las diversas técnicas de procesamiento de las señales, en la figura 52-12 se ilustra un método específico. Las salidas de las dos antenas de cuadro se modulan por medio de las componentes individuales de una onda de forma cuadrada de 32 Hz, defasadas 90° (+90° o -90°, dependiendo de si la antena está montada encima o debajo del fuselaje.) Estas dos señales se suman con una señal escalada y compensada en fase procedente de la antena de sentido. La resultante es una señal modulada en fase por una onda de forma cuadrada de 32 Hz. La fase de su modulación, respecto a la señal de conmutación de 32 Hz original, es proporcional al ángulo de llegada de la señal entrante. Después de la detección, filtrado y limitación de fase, esta señal de 32 Hz se compara en fase con las dos componentes originales de 32 Hz. Esto produce dos voltajes CD, que, después de pasar por una red de conformación no lineal, son proporcionales al seno y coseno del ángulo de llegada. Estas señales se envían a un resolutor (sin-crotrigonometro) de CD, y la dirección de la señal es indicada por un apuntador unido a su eje.

La figura 52-13 ilustra una versión simple de un indicador ADF del tipo de indicador radiomagnético

(RMI). El ángulo de la aguja muestra el rumbo de la estación respecto a la nariz de la aeronave. El rumbo real a partir del norte magnético del NDB de la posición de la aeronave está dado por el valor de la carátula de la brújula señalado por la aguja. En todas las aeronaves excepto las ligeras, la indicación ADF suele combinarse con la de otros auxiliares de la navegación, tales como el rumbo de VOR o TACAN, y se presentan en la visualización del indicador de situación horizontal (HSI) (fig. 52-6).

Inevitablemente, la estructura de la aeronave distorsiona los diagramas de las antenas de cuadro y de sentido. Esto produce errores importantes hasta de 20° a ciertos ángulos. Estos "errores de cuadrante" se compensan en cada tipo de aeronave específico por medio de los circuitos de conformación del voltaje del receptor ADF. La RTCA establece normas de rendimiento mínimo para receptores ADF de avión<sup>16</sup>. Un receptor ADF de aeronaves ligeras pesa alrededor de 3 kg, ocupa 2 500 cm<sup>3</sup> y consume menos de 10 W. Para una unidad de línea aérea, las cifras correspondientes serían 5 kg, 3 800 cm<sup>3</sup> y 25 W.

**Parámetro y exactitud del sistema.** Por su capacidad de operar en las bandas LF y MF, el sistema NDB/ADF no tiene limitaciones de trayectoria óptica. Sin embargo, para el uso aeronáutico la cobertura utilizable se considera situada en cualquier punto en que la razón de la onda terrestre polarizada verticalmente sobre la potencia total de la onda reflejada sea de 10 dB o mayor. La ICAO recomienda que dentro del área nomi-

nal de cobertura del NDB, la intensidad de la señal no sea menor de 70  $\mu\text{V/m}$ , salvo en regiones tropicales, donde se prefiere una intensidad de 120  $\mu\text{V/m}$ . Los requisitos específicos dependen de factores tales como la conductividad de la superficie y el ruido de fondo de las áreas urbanas.

En el anexo C de la referencia 4 se presenta un análisis detallado de estos factores.

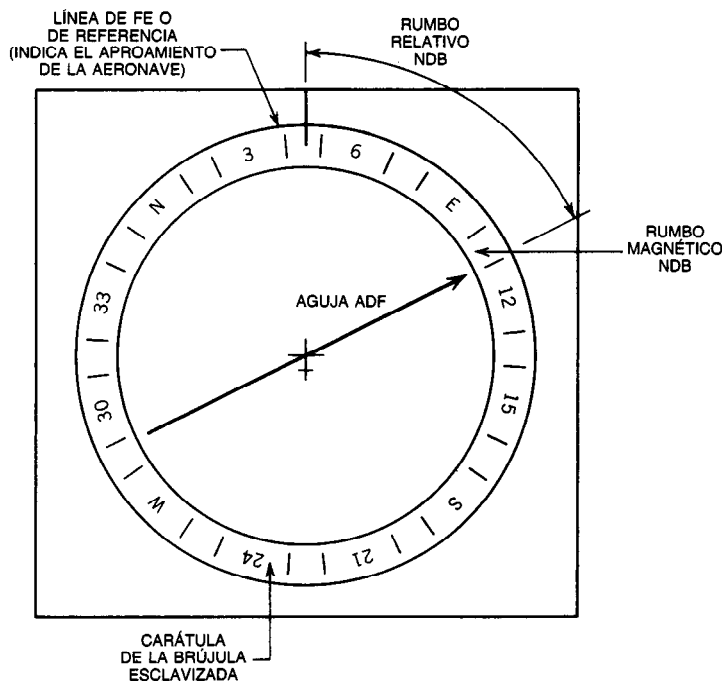
En Estados Unidos se definen tres clases de NDB para el servicio en ruta:

Código de clase	Potencia (W)	Radio utilizable (nmi)
MH	Menos de 50	25
H	50 a 1 999	50
HH	2 000 o más	75

La exactitud de rumbo del equipo de avión debidamente calibrado es del orden de  $\pm 3^\circ$ . Dentro del radio utilizable del NDB, los errores a corto plazo debidos a efectos topográficos y de propagación no deben exceder de  $\pm 10^\circ$ , según exigencias de la FAA. Por la noche, durante periodos de propagación anómala y en particular en la vecindad de tormentas magnéticas, es probable que la exactitud del sistema NDB/ADF se degrade en forma considerable.

#### Navegación por áreas

El sistema actual de vías aéreas sobre tierra de cualquier país se define por medio de líneas rectas que



**Fig. 52-13.** Indicador del ADF (radiogoniómetro automático). NDB = radiofaro no direccional.

unen los auxiliares de la navegación de fuente puntual del tipo rho-theta (VOR/DME, NDB). En muchos casos esta estructura rígida introduce desviaciones innecesarias en la ruta comprendida entre dos puntos. En años recientes se ha puesto a disposición una serie de auxiliares de cálculo a bordo, conocida como sistemas de navegación por áreas (R-NAV). Éstos están disponibles ahora incluso para las aeronaves livianas de un solo motor. Mediante su uso, un piloto que se encuentre dentro de la cobertura de dos o más estaciones VOR/DME, puede definir sitios VOR/DME "fantasmas" o "puntos a medio camino" a lo largo de la ruta de vuelo deseada. Después, el piloto vuela directamente entre estos puntos de la ruta mediante el uso de la instrumentación de vuelo estándar. En Estados Unidos se han definido y representado en mapas varias rutas de cobertura de áreas especiales (R-NAV) para su empleo por aeronaves en condiciones de vuelo por instrumentos. No obstante, no se disponen para muchas áreas de alta densidad de tráfico. Esto se debe principalmente a la incapacidad del sistema de control del tráfico aéreo actual, estructurado como está para guiar una aeronave que vuela en rutas aéreas fijas de un punto a otro, de controlar estas rutas R-NAV "aleatorias".

Muchas aeronaves de mayor tamaño (B-747, L-1011, DC-10) ahora cuentan con sistemas de navegación inerciales como equipo estándar. Éstas pueden seguir las rutas R-NAV definidas —cuando están disponibles— mediante el uso del sistema inercial para el guiage, si bien deben portar a bordo equipo VOR/DME como su sistema de navegación principal.

### Radiobaliza marcadora

Las radiobalizas marcadoras o radiofaros de aproximación son transmisores de baja potencia acoplados a antenas de haz estrecho orientadas verticalmente. Éstos se instalan en sitios donde se desea dar a una aeronave una indicación precisa de su paso por ese punto. Todas las radiobalizas marcadoras operan, en todo el mundo, a una frecuencia de 75.0 MHz, con base en la baja potencia y directividad de la antena para la separación de las señales.

Las radiobalizas marcadoras se implementaron originalmente como un medio de identificación del centro y de los sectores de rumbo del ahora obsoleto radiofaro direccional de cuatro rumbos (ref. 2, cap. 2). Como auxiliares de vuelo en ruta, todavía se pueden encontrar en la definición de puntos específicos ("intersecciones") a lo largo de vías aéreas definidas entre instalaciones NDB donde, desde luego, no se dispone de información DME para definir la distancia a partir de la instalación. En la actualidad, el uso más importante de las radiobalizas marcadoras es como parte del sistema de aterrizaje por instrumentos (ILS) (cap. 53).

La ICAO define las características de dos tipos de radiobalizas marcadoras en ruta (ref. 4, sec. 3.6). Las radiobalizas marcadoras Z se utilizan para identificar un punto específico y tienen un diagrama (patrón) de antena en forma de cono vertical. Las radiobalizas de abanico, por otra parte, definen una posición a lo largo de una vía aérea y tienen un haz elíptico, don-

de el eje menor está situado a lo largo de la vía aérea y el eje mayor es perpendicular a la ruta. La frecuencia portadora de 75 MHz se modula en amplitud a un valor nominal del 95% con un tono de 3 000 Hz, que se manipula encendiéndolo y apagándolo para producir señales de identificación en clave Morse. El área definida por la radiobaliza es función de la potencia radiada transmitida, sensibilidad del receptor de la nave y altitud de ésta. La intensidad nominal de señal para la detección se especifica en 1.5 mV/m. Con sensibilidades del receptor ajustadas en forma correcta, la radiobaliza Z genera esta intensidad sobre un área con diámetro aproximado de 1.5 nmi para una aeronave que está situada 3 000 pies por encima de la estación. Una radiobaliza de abanico define un área aproximada de  $2 \times 10$  nmi a 3 000 pies, y de  $18 \times 6$  nmi a 13 000 pies por encima de la estación.

**Equipo de tierra.** Los transmisores de las radiobalizas Z tienen aproximadamente 4 W de potencia. La antena es un "torniquete" de cuatro elementos de media onda horizontales colocados un cuarto de longitud de onda sobre un contrapeso. El ajuste de fase correcto de los elementos genera el haz cónico vertical que se requiere, con polarización circular horizontal. La potencia del transmisor de la radiobaliza de abanico es comúnmente de 100 W. Un tipo de arreglo de antenas consta de cuatro elementos de media onda colineales, hacia el centro, con secciones equilibradoras entre los elementos. Toda la pieza está situada un cuarto de longitud de onda sobre el contrapeso. Esto produce un haz elíptico vertical, donde la dimensión estrecha es perpendicular a la línea de elementos de la antena.

**Equipo de a bordo.** Las aeronaves livianas utilizan una antena de alambre de media onda simple montada sobre separadores a unos 8 cm por debajo del fuselaje, paralelos a la línea central. Las aeronaves de mayor tamaño y alta velocidad utilizan antenas de cavidad resonante montadas al ras o casi al ras con la parte inferior del fuselaje. El receptor de la aeronave es una unidad controlada por cristal a la frecuencia fija de 75 MHz. No es deseable una sensibilidad mayor de 1.5 mV/m. Cuando la señal excede el umbral, las señales de identificación de 3 000 Hz detectadas se conmutan al sistema de audio de la nave. El voltaje detectado en un circuito resonante de 3 000 Hz se utiliza para activar una luz de señal blanca visible en el tablero de la cabina del piloto, que destella cuando se enciende y apaga la modulación de 3 000 Hz.

### DECCA

El sistema DECCA, que ahora es un auxiliar de la navegación marítima principalmente (descrito en la sección 52.3), se utiliza asimismo en aeronaves de ala fija y giratoria. En el pasado, el sistema fue propuesto a la ICAO para ser adoptado como el principal sistema auxiliar de la navegación de aeronaves de corta y mediana distancia en todo el mundo. Sin embargo, en 1959 la ICAO eligió el sistema VOR/DME por una diversidad de factores políticos y técnicos, uno de los cuales era la dificultad que se tenía en ese tiempo de producir

un receptor y una pantalla DECCA compactos, pero con un costo razonable para aeronaves ligeras.

Como sistema auxiliar para la navegación de aviones, el sistema DECCA ahora se utiliza principalmente en helicópteros. La cobertura de área de DECCA es particularmente adecuada para las rutas aleatorias de estas aeronaves. Además, como DECCA no está limitado por efectos de alcance óptico y tiene exactitudes del orden de decenas de metros en la mayor parte de las áreas de cobertura utilizables, el sistema es de gran valía para ayudar a los helicópteros a hacer descensos en sitios de aterrizaje no preparados y sin asistencia de otros sistemas auxiliares para la navegación. En el Mar del Norte, el DECCA se utiliza intensamente en los helicópteros que dan servicio a las numerosas plataformas de extracción de petróleo del área.

En Europa, donde la cobertura de DECCA es casi completa, el sistema de recepción está instalado en aviones de British Airways para servicios nacionales y en Europa, y se emplea en muchas aeronaves corporativas de ala fija. Los helicópteros y algunas aeronaves de la British Army y Royal Air Force están equipados con receptores DECCA.

En los receptores DECCA de a bordo<sup>17</sup> no existe multiplicación de las frecuencias componentes que se reciben, sino que, en cambio, todas las componentes se dividen (entre 5, 6, 8 o 9) a la frecuencia común  $f$ , para la cadena DECCA en uso. La posición se mide como fracciones (comúnmente 1/1024) de "zonas", en vez de en "calles" o "pasillos" como en los receptores marinos. Las ambigüedades dentro de la zona ocasionadas por el proceso de división se resuelven por medio de señales de identificación de "calles" de pulsos múltiples. La combinación de estos procedimientos permite hacer comparaciones de fase directas en  $f$  y también en  $0.2f$ . De esta forma se pueden resolver las ambigüedades automáticamente dentro de una zona y dentro de grupos de cinco zonas.

El elemento más importante de DECCA para la navegación de aeronaves fue el desarrollo del mapa de navegación en movimiento o diario de vuelo (cuaderno de bitácora). En su forma básica, en este registro se emplean los conteos y fracciones de zona para los dos "colores" que se utilizan como las coordenadas rectangulares  $X - Y$  en la representación de un mapa arrollado que avanza. Como las coordenadas originales son hiperbólicas, esto produce algunas veces una imagen muy distorsionada cuando se emplean mapas elaborados en forma especial. Sin embargo, en la práctica esto no afecta de manera importante la capacidad del piloto de volar por una pista pretrazada o de identificar puntos de referencia. En una segunda forma de bitácora de vuelo DECCA se utiliza una computadora digital para realizar una conversión de coordenadas de hiperbólicas a rectangulares, lo que hace posible utilizar mapas con distorsión considerablemente reducida.

Son posibles muchas combinaciones de equipo, pero una instalación DECCA moderna (Mk 19) de a bordo podría constar de un receptor y amplificador de antena, la computadora digital, la representación del diario de vuelo, un medidor de identificación de zona

(para identificar grandes errores) y dos unidades de control. Este sistema pesa aproximadamente 19 kg y consume unos 240 W. La información del mapa se codifica en una pista óptica binaria de nueve bits que está a un costado del rollo plástico del mapa. La salida se envía a la computadora asociada para producir un control de ciclo cerrado de la posición en el mapa. Una pista similar rige la posición izquierda y derecha de un cursor.

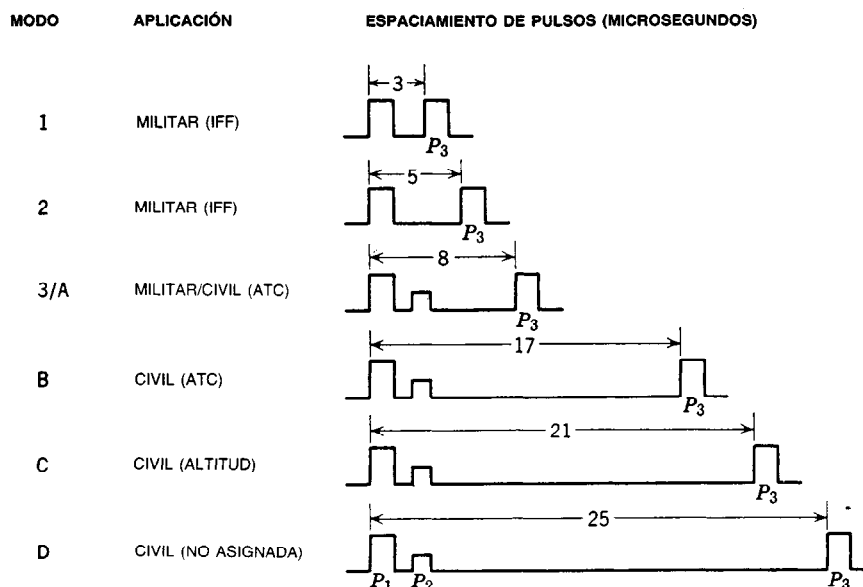
#### Radar secundario de vigilancia/ATCRBS

El radar primario experimenta varios problemas cuando es utilizado como herramienta de vigilancia por el control del tráfico aéreo en tierra (ATC). Las señales se atenúan por precipitación, la sección transversal para el radar que ofrece la aeronave ligera es demasiado pequeña para generar respuestas (ecos) consistentes de radar y, lo que es más importante, el radar primario no puede identificar directamente si una respuesta proviene de una aeronave específica. El radar secundario de vigilancia (SSR) elimina estos problemas casi por completo. En Estados Unidos, el SSR se conoce como sistema de radiobaliza para el informe por radar del control del tráfico aéreo (ATCRBS).

Creado a partir del sistema Mark X IFF (identificación de amigo o enemigo) de la Segunda Guerra Mundial, el SSR se basa en un radiofaro de respuesta (transpondedor) portado en cada aeronave, el cual da respuestas cuando es interrogado por señales codificadas de la estación de tierra. Mediante el uso de una antena direccional y midiendo el retardo del viaje de ida y vuelta, la terminal de tierra mide el rumbo y distancia a cada objetivo. En el estado actual de desarrollo, cada aeronave equipada puede enviar de regreso un código de identificación único además de la lectura de su altitud barométrica. Ahora se están desarrollando extensiones de este sistema en Estados Unidos; el sistema de baliza de transmisión de datos entre tierra y aire o sistema de direccionamiento discreto (DABS) y el sistema de evasión de colisiones por radiobaliza (B-CAS). El DABS hace posible el intercambio de mensajes limitados de control del tráfico aéreo entre la tierra y la aeronave específica; el B-CAS ofrece una jerarquía de facultades de detección de aeronaves para evitar colisiones actuando bien en forma autónoma, o bajo supervisión de la estación de tierra.

**Descripción del sistema básico.** Los parámetros del sistema están definidos por especificaciones de Estados Unidos<sup>18</sup> y la ICAO (ref. 4, sec. 3.8). Todas las transmisiones de tierra a aire se hacen a una frecuencia de 1 030 MHz. La frecuencia de respuesta en todos los casos es de 1 090 MHz. Las señales de interrogación y respuesta constan de grupos de pulsos codificados por el espaciamiento de los pulsos. Cada pulso de interrogación tiene una amplitud nominal de  $0.8 \pm 0.1 \mu s$ , y cada uno de los pulsos de respuesta tiene  $0.45 \pm 0.1 \mu s$  de ancho. Los tiempos de ascenso y descenso controlados se utilizan para minimizar la ocupancia espectral.

Los seis modos de interrogación militar y civil se ilustran en la figura 52-14. Los transpondedores de aeronaves civiles están diseñados en general para re-



**Fig. 52-14.** Modos de pulsos de interrogación que utiliza el radar secundario de vigilancia. IFF = identificación de amigo o enemigo; ATC = control del tráfico aéreo. Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.

conocer únicamente los modos 3/A y C. A las interrogaciones de modo 3/A la aeronave responde con su código asignado en forma única; al modo C responde con un código que indica la altitud barométrica. La antena de tierra que se utiliza para radiar  $P_1$  y  $P_3$  en la figura 52-14 tiene un estrecho haz de acimut (comúnmente de  $3^\circ$ ). Por lo general esta antena está montada en, y gira con, la antena del radar primario ATC con que está asociada la instalación. Esto hace posible la correlación entre las respuestas del radar primario y el secundario. El ritmo de interrogación depende de la velocidad de rotación de la antena; por lo general se fija de modo que se obtengan de 4 a 5 respuestas sucesivas de cada aeronave en el ancho del haz de  $3^\circ$ , hasta un máximo de 450 por segundo. Una secuencia de código de interrogación civil típica es dos veces el modo 3/A, seguido de una vez el modo C.

En los modos civiles se transmite otro pulso,  $P_2$ , a fin de suprimir las respuestas de la aeronave a interrogaciones de lóbulo lateral o interrogaciones demoradas de trayectoria múltiple.  $P_2$  sigue a  $P_1$  en  $2 \mu s$  y se radia empleando una antena omnidireccional tal que su potencia radiada efectiva (ERP) sea menor que la de  $P_1$  y  $P_3$  en el haz principal de su antena, pero mayor que en los lóbulos laterales y posteriores. La radiobaliza de respuesta (transpondedor) del avión responderá sólo si se cumplen las siguientes condiciones:

1. El espaciamiento de  $P_1$  a  $P_3$  es correcto dentro de  $\pm 1 \mu s$  en el modo seleccionado.
2. La amplitud de  $P_3$  es igual a la amplitud de  $P_1$  (+3 dB, -1 dB).
3. La amplitud de  $P_2$  está más de 9 dB por debajo de la amplitud de  $P_1$ .

La respuesta del transpondedor consta hasta de 15 pulsos. Toda respuesta contiene los dos pulsos de encuadre, separados  $20.3 \mu s$ . La respuesta de identificación a las interrogaciones de modo 3/A contiene uno de 4 096 códigos discretos (del 0000 al 7 777 en octal), seleccionado manualmente por el piloto en respuesta a la instrucción ATC (que por lo general se fija una sola vez, antes del despegue). El valor del código se envía por presencia o ausencia de los pulsos A, B, C, D según sus valores aritméticos, como se indica en la figura 52-15; por ejemplo, el código 4 034 =  $C_1 + C_2 + A_4 + D_4$ . Se asignan ciertos códigos fijos en circunstancias especiales, por ejemplo 7 700 = urgencia, 7 600 = falla en las comunicaciones y 7 500 = asalto. Para transmitir datos de altitud, la aeronave requiere un altímetro de codificación que genere datos binarios en un código especial de 11 bits. Los 11 bits se conmutan a los codificadores de pulsos A, B, C, D del radiofaro de respuesta ( $D_1$  no se utiliza) cuando se responde a interrogaciones de modo C. La referencia de presión para el altímetro de codificación siempre está fija en 29.92 pulg de columna de Hg (1013.2 mb). Las correcciones de la diferencia entre ésta y la presión de referencia local real se hacen en el equipo de tierra tras decodificar la respuesta.

Para identificar positivamente una aeronave y para proteger contra equivocaciones al establecer el código de identificación, se utiliza el pulso de identificación de posición especial de la figura 52-15. Éste se transmite con todas las respuestas durante unos 30 segundos después de que el piloto oprime un botón "ident" cuando la estación de control del tráfico aéreo le pide hacer esto. Este pulso genera una marca especial en la pantalla del controlador aéreo que trabaja con esa aeronave.

**Procesamiento y visualización en tierra.** Debido al uso de un par de frecuencia común (1 030 y 1 090 MHz), una aeronave responderá a todas las interrogaciones dentro de su trayectoria óptica. La interferencia resultante en la pantalla de una estación de tierra, provocada por respuestas de una aeronave en su área de cobertura a interrogaciones de una estación de tierra diferente, se conoce como respuesta (o señal) no deseada. Como esta respuesta no está sincronizada con las interrogaciones de la estación de tierra primaria, se elimina al aceptar en la pantalla sólo aquellas respuestas que se correlacionen en distancia en dos o más interrogaciones sucesivas. La "mutilación" es una interferencia producida por dos aeronaves con distancias reales y acimuts similares, que responden a la misma interrogación y que producen respuestas que se superponen. Aunque este tipo de error es difícil de eliminar, es relativamente fácil de detectar.

En la figura 52-16 se ilustra una típica pantalla de controlador asociada con la etapa A del sistema del espacio aéreo nacional (NAS) de Estados Unidos<sup>19</sup>, que se implementa en todos los centros de control del tráfico de rutas aéreas (ARTCC) que manejan tráfico en ruta. Es posible el seguimiento de "radar" primario, "radar" (radiobaliza) secundario (SSR) y objetivos correlacionados, con etiquetado alfanumérico de la identidad y altitud (en cientos de pies) de cada aeronave objetivo en relación con las respuestas de SSR. En áreas terminales se ha instalado un sistema Sperry-Univac conocido como sistema terminal automatizado de radar III (ARTS-III) en 64 aeropuertos de Estados Unidos. Este sistema tiene una capacidad de despliegue similar, más la capacidad de medición de la velocidad respecto a tierra, advertencia de baja altura, extrapolación de seguimiento durante la pérdida o mutilación de señales, y predicción de conflictos limitada. Muchas de las agencias para el espacio aéreo nacionales de todo el mundo, han instalado medios de procesamiento en curso y terminal y de visualización similares.

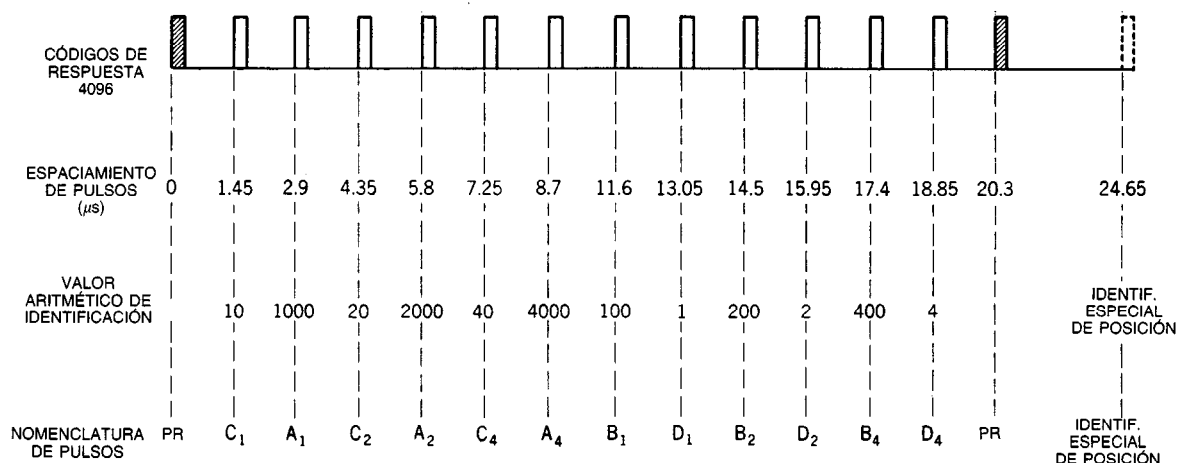
**Equipo de a bordo.** El transpondedor de a bordo es una unidad de tamaño compacto y, en esta época de microcircuitos integrados, relativamente simple. Las frecuencias fijas de transmisión y recepción simplifican el diseño RF. Se utilizan antenas verticalmente de segmento en cuarto de onda, con un diagrama de acimut efectivamente horizontal. En la cabina, la interfaz del piloto consta de cuatro interruptores accionados por el pulgar para seleccionar el código asignado más el botón "ident" para transmitir el pulso especial de identificación de posición. En muchas instalaciones, una luz "de respuesta" en el tablero destella en respuesta a interrogaciones exitosas de tierra. La RTCA establece normas de rendimiento mínimo de los transpondedores SSR del avión<sup>20</sup>.

Los parámetros del sistema SSR son los siguientes:

Alcance operacional (del interrogador)	200 nmi
Altitud operacional	100 000 pies
ERP máxima del interrogador ( $P_1$ y $P_3$ )	52.5 dBW (500 W a 1.5 kW de potencia real del transmisor)
Nivel de activación del receptor de la aeronave (mínimo para el 90% de probabilidad de respuesta)	-101 dBW
Potencia pulsátil pico del transpondedor de la aeronave	21 a 27 dBW (125 a 500 W)
Sensibilidad del receptor de tierra	-115 dBW

#### Sistema de transmisión de datos entre tierra y aire (DABS)

El sistema DABS<sup>21,22</sup> (sistema de transmisión de datos entre tierra y aire o sistema de radiobaliza de direccionamiento discreto) está siendo desarrollado para la



**Fig. 52-15.** Códigos de respuesta del transpondedor que se utilizan en el radar secundario de vigilancia. PR = pulso de referencia o de encuadre; IDENTIF. = identificación. Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.



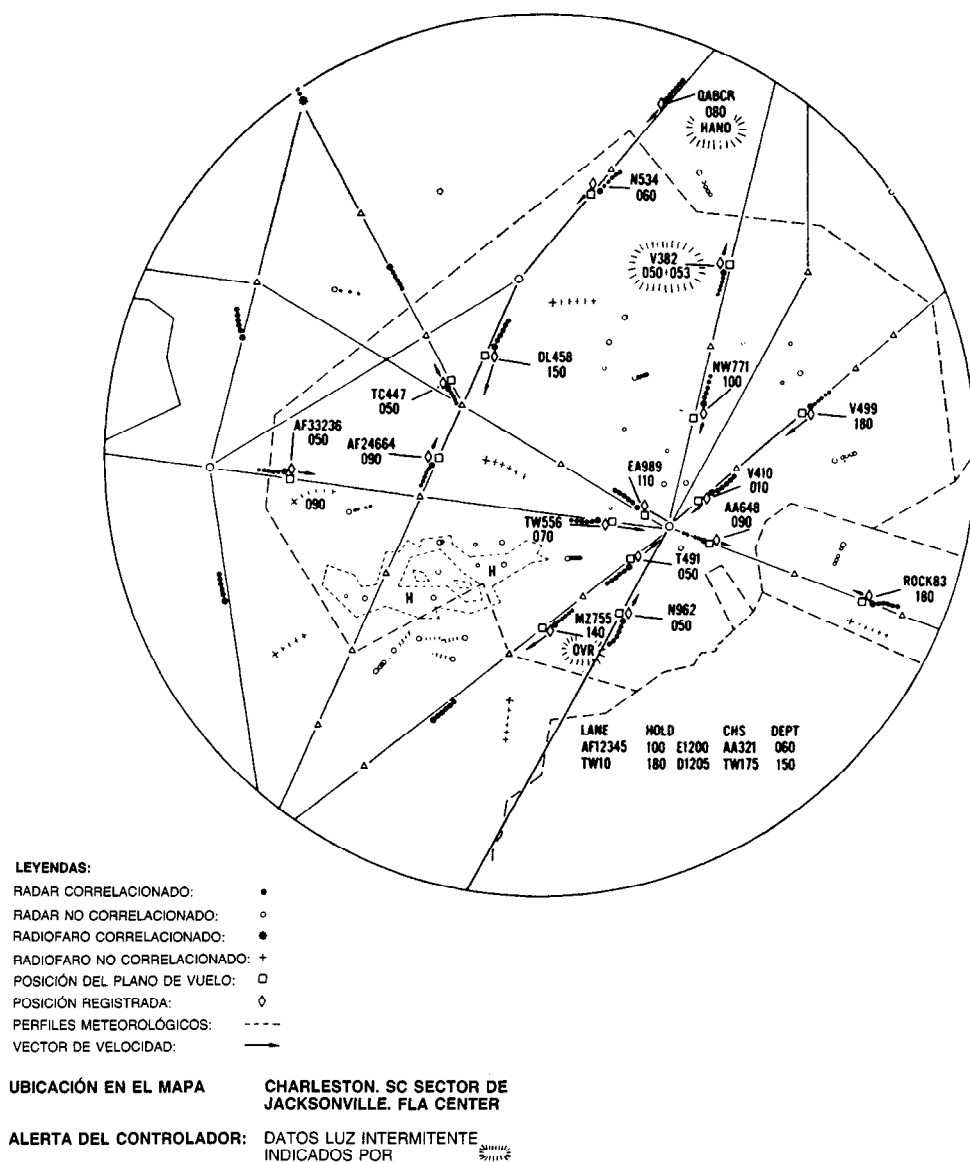


Fig. 52-16. Ejemplo de despliegue pictórico automático del sistema aeroespacial nacional. Tomado de Kayton y Fried, dirs. eds.<sup>12</sup>, con permiso.

FAA por los Lincoln Laboratories, del Massachusetts Institute of Technology. Es resultado de la evolución de ATCRBS, y en él se incorpora una capacidad de enlace de datos de dos vías para transmitir mensajes de tierra a aire y vuelta a tierra. En el ATCRBS, la interrogación es espacial; todas las aeronaves presentes en el campo del haz de la antena de tierra responden. El DABS asigna a cada aeronave un código único, y tiene la capacidad de elegir respuestas sólo de la aeronave seleccionada. Por tanto, tiene asimismo la capacidad de programar interrogaciones con base en la distancia predicha de la aeronave, con lo cual se reduce la interferencia provocada por la superposición de respuestas

(mutilación) y se incrementa considerablemente la capacidad del sistema respecto a la del ATCRBS. La capacidad de direccionamiento discreto elimina asimismo el problema de las respuestas no deseadas a interrogaciones distantes. Como la capacidad de vigilancia mejorada, la función de enlace de datos del DABS ofrece la capacidad de comunicación para el servicio de asesoramiento y resolución de tráfico automáticos (ATARS) del control del tráfico aéreo. También sirve como elemento principal del sistema de evasión de colisiones por radiobalizas (B-CAS) propuesto.

El DABS utiliza las mismas frecuencias y casi los mismos formatos de señales que el ATCRBS. Las

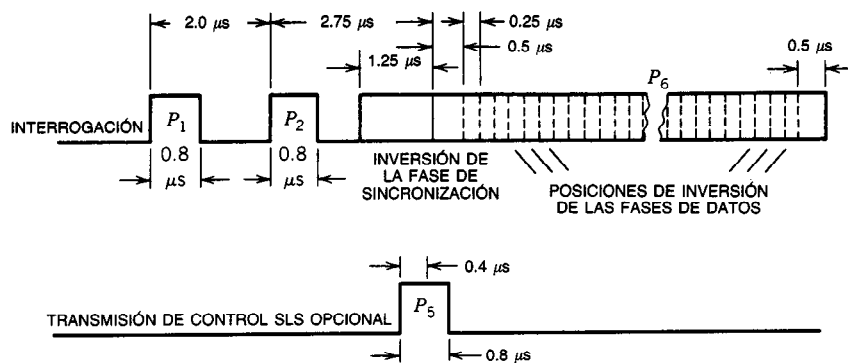


Fig. 52-17. Interrogación DABS (sistema de transmisión de datos entre tierra y aire o sistema de direccionamiento discreto). SLS = supresión del lóbulo lateral. De Orlando y Drouilhet<sup>22</sup>, con autorización.

aeronaves equipadas con DABS y ATCRBS operarán dentro de una estructura común hasta que todas las aeronaves con ATCRBS se mejoren con el equipo DABS. Las especificaciones de la siguiente generación de equipo de transpondedores SSR de líneas aéreas incluyen la capacidad de uso del ATCRBS y DABS<sup>23</sup>.

Se utilizan cinco tipos de señales en el sistema DABS.

1. **Interrogación de todas las llamadas de ATCRBS/DABS.** Todos los transpondedores ATCRBS y DABS en el campo del haz interrogador responden con datos de vigilancia (modo A o modo C, según se solicite).
2. **Interrogación de todas las llamadas sólo de ATCRBS.** Igual que en (1), pero la interrogación es ignorada por las aeronaves con DABS.
3. **Respuesta ATCRBS.** En respuesta a (1) para ATCRBS y DABS y a (2) para ATCRBS.
4. **Interrogación sólo de DABS.** Se utiliza para solicitar datos de vigilancia de una aeronave específica o de todas las equipadas con DABS en el campo del haz de la antena, o para enviar o solicitar mensajes de comunicaciones.
5. **Respuesta DABS.**

Existen algunas variaciones sobre estos puntos. Por ejemplo, a ciertas aeronaves equipadas con DABS se les debe ordenar ignorar las interrogaciones "llamada general" (cierre forzoso), y para comunicaciones se han definido hasta ahora tres tipos de mensajes.

La interrogación de "llamada general" es la misma que una interrogación ATCRBS (modo A o modo C), salvo que un pulso extra,  $P_4$ , con 1.6 o 0.8  $\mu s$  de longitud sigue 2  $\mu s$  después del pulso  $P_3$ . Los transpondedores ATCRBS responden normalmente e ignoran  $P_4$ . Si  $P_4$  tiene 1.6  $\mu s$  de longitud, los transpondedores DABS responden asimismo con una respuesta de tipo ATCRBS. No obstante, todos los transpondedores DABS están diseñados para ignorar la interrogación cuando  $P_4$  tiene sólo 0.8  $\mu s$  de longitud.

Una interrogación sólo de DABS tiene la forma que se muestra en la figura 52-17. Consta de un preámbulo

de dos pulsos ( $P_1, P_2$ ), seguido de un bloque de datos ( $P_6$ ) que contiene una señal de sincronización. Dentro del bloque de datos se transmite la información a razón de 4 megabits por segundo empleando la manipulación por defasamiento diferencial binaria (DPSK). El preámbulo  $P_1$  y  $P_2$ , que se transmite en el haz interrogador principal, sirve para suprimir los transpondedores ATCRBS, ya que ellos interpretarán los términos  $P_2$  como recibidos en el mismo nivel que  $P_1$ , lo que indica una interrogación de lóbulo lateral. Después de 1.25  $\mu s$  del inicio del bloque de datos DABS, se produce siempre una inversión de fase de 180°. Esta señal de sincronización debe ser reconocida por el transpondedor DABS a fin de dar inicio al procesamiento de datos subsiguientes. El pulso de supresión del lóbulo lateral (SLS),  $P_5$ , es transmitido por el interrogador DABS mediante una antena omnidireccional. Si el transpondedor no está en el lóbulo principal del interrogador de haz estrecho, el nivel más alto recibido de  $P_5$  encubrirá la inversión de la fase de sincronización y hará que se suprima la respuesta.

Cada bloque de datos tiene una longitud de 56 o 112 bits. De éstos, 24 bits definen la dirección única (en registros) de la aeronave. Ésta sería asignada por un número de cola permanente o, para el uso de las líneas aéreas, por un número para la compañía y número de vuelo. Una aeronave DABS sólo responde cuando reconoce su dirección en la interrogación, salvo que se utilice un código "todo unos" para la dirección DABS "llamada general". El formato de los mensajes se define en la referencia 21. Se toman medidas previendo la transmisión de mensajes de longitud extendida (ELM), en cuyo caso se envían bloques múltiples a la misma aeronave antes de que se genere una respuesta. El formato de señales de DABS ha sido adoptado por la ICAO y se denomina SSR de modo S.

El formato de la respuesta DABS se ilustra en la figura 52-18. Éste vuelve a constar de un preámbulo más un bloque de datos de 56 o 112 bits. Sin embargo, ahora se utiliza la modulación de la posición del pulso para datos a una velocidad de 1 Mb/s empleando pulsos de 0.5  $\mu s$ . El preámbulo de cuatro pulsos hace posible que la respuesta de DABS se pueda distinguir sin

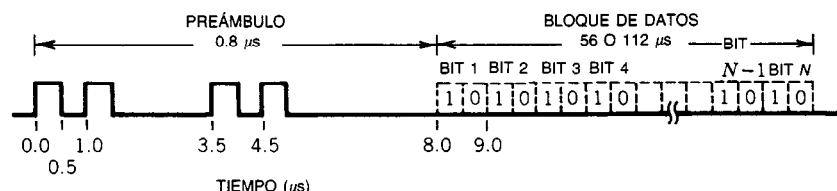


Fig. 52-18. Respuesta DABS (sistema de transmisión de datos entre tierra y aire). De Orlando y Drouilhet<sup>22</sup>, con permiso.

dificultad de cualesquiera respuestas de ATCRBS superpuestas. Se incluye un código de verificación de paridad de 24 bits en los mensajes de enlace ascendente y descendente de DABS para la detección y corrección de errores.

Las características de funcionamiento de DABS se resumen en la tabla 52-2. La cifra de acimut  $\sigma$  que aparece en ella se obtiene mediante el procesamiento monopolso en antena de las respuestas del interrogador de tierra. Varios interrogadores ATCRBS en operación están dotados de antenas adecuadas para el procesamiento monopolso. Desde 1980 han estado en uso unidades experimentales DABS a bordo, y se han probado con interrogadores experimentales DABS y en un entorno de ATCRBS estándar. Las instalaciones operacionales DABS utilizarán antenas de avión dobles (encima y debajo del fuselaje) con combinación de receptores en diversidad, a fin de mejorar la confiabilidad de la transferencia de mensajes.

**Servicio de asesoramiento y resolución de tráfico aéreo automáticos (ATARS).** Como parte del sistema DABS futuro, el ATARS está planificado para ser un sistema auxiliar de evasión de colisiones con base en tierra orientado al piloto. En cada sitio interrogador DABS, el ATARS genera archivos de historial de seguimiento de todas las aeronaves en cobertura (objetivos equipados con DABS, equipados con ATCRBS y de radar primario equipados sin transpondedor). En el caso de aeronaves equipadas con DABS y equipadas con capacidad de codificación de altitud ATCRBS (modo C), se trata de un seguimiento tridimensional. Mediante el uso de la capacidad de enlace de datos de tierra a aire, el ATARS transmite a aeronaves equipadas con DABS asesoramiento sobre la presencia de otras aeronaves registradas y seguidas en su vecindad (en el caso del tráfico de modo no C y sólo con radar primario, esto sólo se puede hacer en el plano horizontal). Cuando los algoritmos de ATARS detectan un conflicto potencial, el enlace de datos envía recomendaciones de maniobra (virar a la izquierda, virar a la derecha, ascender, descender) a las aeronaves adecuadas equipadas con DABS. El ATARS envía asimismo mensajes de alerta a la instalación responsable del control del tráfico aéreo siempre que se transmite una recomendación.

**Sistema de evasión de colisiones por radiobaliza (B-CAS).** El B-CAS ofrece a las aeronaves equipadas la capacidad autónoma de detectar (y maniobrar en

correspondencia para evitar la colisión) otras aeronaves equipadas con DABS y ATCRBS de modo C cuando están fuera de la cobertura de ATARS<sup>24</sup>. Éste es otro adelanto evolutivo del DABS, para el cual ya se está probando equipo experimental en áreas de tráfico de alta densidad.

Una aeronave DABS con capacidad B-CAS interroga periódicamente a todas las aeronaves equipadas con DABS o ATCRBS en modo C en su vecindad, y mide las diferencias de distancia y altura entre dos aeronaves en el aire a partir de sus respuestas. Se conservan archivos de control de seguimiento a bordo para cada aeronave que se considere representa un conflicto potencial. Para ayudar a la adquisición, todas las aeronaves DABS emiten periódicamente en forma espontánea una transmisión de identificación. Entre aeronaves equipadas con DABS, el enlace de datos de comunicaciones entre dos estaciones en el aire hace posible el intercambio de recursos de velocidad y altura de las aeronaves para evitar la reinterrogación continua de aeronaves distantes que no representen una amenaza. No existe una medición del ángulo de llegada, de modo que la discriminación se debe realizar exclusivamente sobre la base de velocidades de cambio de distancia y altura. Análogamente, las maniobras evasivas de B-CAS se realizan sólo en el plano vertical. Cuando una aeronave B-CAS encuentra una amenaza básica de DABS o ATCRBS, es la aeronave B-CAS la que efectúa la maniobra de evasión si ésta es necesaria. Entre dos aeronaves B-CAS, el enlace de datos entre dos unidades en el aire se utiliza en forma automática para coordinar la maniobra de una o ambas aeronaves.

Se siguen desarrollando planes para la implementación final de ATARS y B-CAS. En otro concepto reciente<sup>25</sup>, el uso de instalaciones con base en tierra pasa a segundo término y la capacidad del avión adquiere la mayor importancia. En este sistema de alerta contra amenazas y evasión de colisiones (T-CAS), las aeronaves completamente equipadas tendrán antenas direccionales de interrogación y respuesta y podrán calcular y realizar maniobras evasivas en los planos horizontal y vertical.

#### Sistema conjunto de distribución de información táctica (JTIDS)

El sistema conjunto de distribución de información táctica (JTIDS) es un sistema de control y mando táctico de escala real en proceso de desarrollo conjunto por las fuerzas armadas de Estados Unidos. Al mismo tiempo que realiza una función de comunicaciones,

**Tabla 52-2. Sistema de transmisión de datos entre tierra y aire (radiobaliza de direccionamiento discreto DABS): resumen de rendimiento**

<b>Vigilancia</b>	
Capacidad	Hasta 700 aeronaves por interrogador
Acimut ( $\sigma$ )	$0.06^\circ \pm 0.033^\circ$ de sesgo
Alcance ( $\sigma$ )	50 pies $\pm$ 150 pies de sesgo
Intervalo de actualización	Cobertura en terminal de 4 s; cobertura en ruta de 5 a 10 s
<b>Enlace de datos</b>	
Capacidad	$\approx 3\%$ del régimen de trabajo para los requisitos totales de mensajes identificados hasta la fecha
Confiabilidad	$> 0.99$ en 4 s para mensajes cortos
Proporción de errores no detectados	$< 10^{-7}$

tiene una importante capacidad navegacional. El JTIDS opera en la misma banda que se utiliza para el TACAN (960 a 1 215 MHz) mediante técnicas de pulsos que hacen posible la realización de operaciones JTIDS y TACAN simultáneas sin que haya una degradación importante debida a interferencia mutua. Aunque la mayor parte de los detalles de funcionamiento son secretos, en las referencias 26 y 27 se presenta una descripción del sistema y su estructura de señales.

Cada usuario (un barco, una estación de tierra o una aeronave) interroga a otros usuarios dentro de la línea de visión y determina la distancia por medio de retardo del viaje de ida y vuelta en los dos sentidos. Cada usuario, cuando se le solicita, envía una estimación de su posición y del error de posición, basada en todos los auxiliares de navegación que el usuario puede estar utilizando. Por tanto, el solicitante original puede hacer una estimación "óptima" de su propia posición mediante el uso de las distancias medidas hasta los otros usuarios y de las estimaciones ponderadas de sus posiciones. Por ejemplo, si una fuente es una unidad de tierra que navega por el sistema de localización global (GPS), esta entrada se ponderaría más que alguna otra fuente, tal como una aeronave que navegara de manera autónoma por radar Doppler.

La resistencia contra interferencia deliberada y la seguridad de los mensajes se obtienen con una combinación de salto de frecuencia y dispersión espectral (de seudoruido) de secuencia directa. Dentro de cada pulso, la información se transmite mediante manipulación continua por defasamiento (CPSK). Se emplean manipulación por desplazamiento de código y codificación en bloques de Reed-Solomon, basadas en un alfabeto de 32 elementos, para reducir la proporción de errores en los mensajes. Como parte de su estructura de señales, el JTIDS incorporará funciones que simulan el TACAN, eliminando en el futuro la necesidad de que aeronaves equipadas con el JTIDS transporten también un equipo TACAN.

### 52.2.3 Sistemas de radionavegación de larga distancia de aeronaves

Las primeras radioayudas que utilizaron las aeronaves para la navegación de larga distancia y oceánica fueron del tipo de baja a media frecuencia, en la banda de 200

a 400 kHz. Prominente entre éstas era el radiofaro direccional de cuatro rumbos<sup>2</sup>, ahora obsoleto, y que se utilizó con éxito en los primeros vuelos realizados entre la costa oeste de Estados Unidos y Hawai en 1927. Este sistema guiaba asimismo las aeronaves a través del Atlántico Norte durante la Segunda Guerra Mundial. Una radioayuda similar en la banda de los 300 kHz era el CONSOL, que se creó durante aquella guerra en Alemania como el SONNE y continuó en operación, principalmente para uso marino. Estas radioayudas se empleaban en forma intermitente para producir "marcaciones" de verificación cruzada en posición durante la navegación, principalmente a la estima. Las radioayudas en estas bandas de frecuencia estaban expuestas al ruido atmosférico producido por la actividad de tormentas eléctricas, graves errores de rumbo por la noche o cuando las señales cruzaban una frontera entre tierra y agua, e interferencia de canal común de estaciones distantes, en especial por la noche.

El radar Doppler de a bordo (sec. 51.8), que opera en la banda de frecuencia de 10 a 20 GHz, proporcionaba inicialmente a las aeronaves de transporte aéreo una ayuda electrónica para la navegación a la estima. En virtud de que no requería estaciones en tierra, este sistema tenía la capacidad de producir datos en cualquier parte del mundo; sin embargo, seguía siendo necesario un medio para tomar puntos fijos (marcaciones) independientes ocasionales. La observación de las estrellas ofrecía esta capacidad en muchos casos. El radar Doppler se ha sustituido casi en su totalidad por la navegación por inercia en el servicio de líneas aéreas oceánicas. No obstante, sigue siendo empleado por muchas aeronaves corporativas de larga distancia y en forma intensa por los militares.

El LORAN-C (que se describe en la sec. 52.3) se utiliza intensamente por las fuerzas militares de Estados Unidos y la OTAN para la navegación en ruta. Los usuarios civiles son en su mayor parte helicópteros que operan sobre el mar lejos de la costa, en particular en el Mar del Norte y la costa estadounidense del Golfo de México para el servicio de equipos petroleros. Es utilizado por algunas aeronaves comerciales de ala fija para la navegación en el Atlántico Norte, donde la cobertura es continua, y también en vuelos entre la costa este de Estados Unidos y regiones del Caribe.

Tabla 52-3. Usuarios de radioayudas de navegación de corta distancia

Sistema	Número proyectado de usuarios (aviones)		
	Civil E.U.A.	Militar E.U.A.	Todo el mundo
OMEGA/VLF	800 <sup>a</sup>	1700 <sup>a</sup>	4000
LORAN C/D	1000 <sup>a</sup>	1100 <sup>a</sup>	4000
Radar Doppler	1000	1500	5000

<sup>a</sup> Tomado de la referencia 1.

La radioayuda de navegación OMEGA (sec. 52.3) se está utilizando cada vez más para la navegación oceánica de aeronaves de líneas aéreas o corporativas<sup>28</sup>. Los receptores OMEGA de aeronaves se están diseñando asimismo para utilizar ciertas señales de comunicaciones VLF muy estables, que se difunden por radio desde diversos puntos en todo el mundo.

En vez de especificar un sistema en particular para la navegación oceánica, la ICAO ha recomendado especificaciones mínimas de funcionamiento en la navegación (MNPS). Estos requisitos MNPS se expresan en términos del error de seguimiento lateral como se indica a continuación:

- Desviación estándar de error (1 $\sigma$ ): 6.3 nm (11.7 km).
- El error excede 30 nmi (55.6 km) por menos de una hora en 1 900 horas de vuelo.
- El error está entre 50 y 70 nmi (92.6 a 129.6 km), menos de una hora en 8 000 horas de vuelo.

La operación con equipo que cumple los requisitos MNPS es ahora obligatoria en ciertas regiones de tráfico de alta densidad, como las rutas "preferidas" que atraviesan el Atlántico Norte.

El número estimado de usuarios de estas radioayudas de navegación de larga distancia se proporciona en la tabla 52-3.

#### Descripciones del equipo

**OMEGA/VLF.** Los principios de operación de OMEGA se describen en la sección 52.3. Al igual que el modo (hiperbólico) de diferencia de fase estándar, los receptores OMEGA de aeronaves suelen estar configurados para trabajar en un modo de telemetría directa, mediante el uso de una referencia de frecuencia estable a bordo. El cambio de fase de cada estación recibida respecto a una posición inicial conocida representa un cambio de distancia para esa estación. La navegación es posible en este modo relativo empleando sólo dos estaciones OMEGA recibidas (modo rho-rho), pero la exactitud se deteriora progresivamente con el tiempo en virtud de la deriva del oscilador de referencia. Con tres estaciones (modo rho-rho-rho), es posible hacer una corrección limitada de esta deriva del reloj de referencia. Además de utilizar los ocho transmisores OMEGA proyectados, muchos receptores de avión están diseñados para operar con transmisores VLF estables en diversas partes del mundo,

tales como los que conserva la marina de Estados Unidos para la transmisión de datos por todo el mundo. Las siete estaciones que se utilizan más comúnmente son:

Estación	Ubicación	Frecuencia (kHz)
NWC	Australia	22.3
NDT	Japón	17.4
GBR	Inglaterra	16.0
NAA	Maine, EUA	17.8
NPM	Hawai	23.4
NSS	Maryland, EUA	21.4
NLK	Washington (est.), EUA	18.6

La potencia de transmisión de estas estaciones es del orden de 1 MW. En una implementación, el equipo de a bordo desactiva la recepción de los canales de OMEGA de 11.33 kHz y superpone las siete señales VLF en una secuencia fija en los segmentos de tiempo que originalmente ocupaban las señales de 11.33 kHz. Por tanto, cada transmisión se muestrea una vez cada 10 segundos. El punto ("fija" o marcación) de posición es un compuesto de señales OMEGA y VLF. Como los transmisores VLF tienen una sola frecuencia con fase arbitraria, no se pueden utilizar para determinar una posición inequívoca, sino sólo en un modo de navegación relativa. Sin embargo, este modo de navegación relativa que da inicio desde una posición conocida, en realidad produce errores de navegación a corto plazo menos graves que el sistema OMEGA puro (navegación hiperbólica). Entre actualizaciones ocasionales de VOR/DME, un sistema OMEGA/VLF relativo que opere en tierra puede funcionar como una radioayuda navegacional de área con errores de posición del orden de 1.5 nmi. En vuelos largos transoceánicos, donde no se dispone de actualización, el error de modo relativo aumenta sin límite. La operación en el modo OMEGA puro comúnmente limitará los errores a más o menos 4 nmi. Para ayudar en el seguimiento y hacer posible la extrapolación en periodos de caída de señales, el equipo OMEGA/VLF de avión suele estar diseñado para aceptar datos de velocidad del aire y aproamiento del sistema de datos del aire de la aeronave.

En la referencia 29 se presentan especificaciones de funcionamiento mínimo para sistemas OMEGA de avión. Algunos parámetros significativos son los siguientes:

Intensidad de campo operacional mínima	10 $\mu$ V/m
Razón mínima de señal sobre ruido (SNR) para adquisición	0 dB en 100 Hz BW
SNR mínima para el seguimiento	-12 dB en 100 Hz BW
Exactitud estática (95% de probabilidad)	$\pm 5$ centipasos (centilanes, CEL) (aproximadamente $\pm 5 \times 0.08$ nmi a 10.2 kHz en la línea base)
Error en la corrección automática de la propagación (95% de confianza)	< 30 centiciclos (CEC) (aprox. 30 $\times 0.16$ nmi en la línea base)
Error de seguimiento dinámico (95% de probabilidad) debido a un giro de 180° a 3°/s o cambio brusco de velocidad de 200 nudos:	< 10 nmi
Tiempo de sincronización inicial (con SNR mínima)	< 6 min

Un sistema típico de a bordo consta de una unidad receptora y procesadora, una unidad de control y pantalla, y una antena. La figura 52-19 ilustra el tablero y las funciones de una unidad comercial de control y pantalla. Al igual que la salida de coordenadas de latitud y longitud, la unidad indica la distancia que falta por recorrer y el error transversal de seguimiento en un rumbo entre dos puntos seleccionados del camino. Las señales de guiage a la izquierda y a la derecha se proporcionan a un medidor tipo indicador de la desviación del rumbo (CDI) o un piloto automático.

Las antenas de cuadro con diagrama de campo H son las preferidas para la recepción de OMEGA. Un alojamiento de antena común tiene unas dimensiones de 16.5  $\times$  35.5 cm y sobresale 4.75 cm de la superficie de montaje. Sin embargo, las antenas de cuadro son muy susceptibles a la interferencia de 400 Hz causada por los sistemas eléctricos CA de las aeronaves, de manera que es necesario localizar cuidadosamente dichos sistemas en la superficie de cada tipo de aeronave para encontrar la ubicación más neutra eléctricamente. También se utilizan antenas de campo E del tipo capacitivo de cavidad. En este caso se debe tener cuidado de minimizar los efectos de la estática de la precipitación. En algunas instalaciones de aeronaves, el uso de la antena de sentido del sistema radiogoniómetro automático (ADF) ha resultado satisfactorio para la recepción del OMEGA.

**LORAN-C.** Las características del LORAN-C se presentan en la sección 52.3. Los receptores LORAN-C de aeronaves suelen requerir un control mínimo del operador, después de la selección de la cadena. La adquisición y sincronización con el tercer ciclo del pulso de la señal de ondas terrestres se realizan en forma automática. A largas distancias, la operación con ondas ionosféricas (indirectas), con exactitud degradada, se puede seleccionar en forma manual. La RTCA define las normas de funcionamiento recomendadas para el equipo receptor LORAN de aviones (ref. 30). Algunas de éstas se tabulan a continuación:

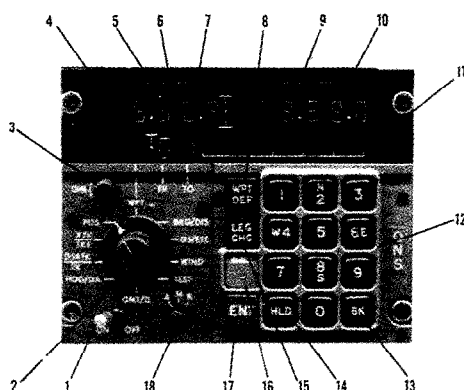
Alcance de la señal de operación	30 dB por encima de 1 $\mu$ V/m a 120 dB por encima de 1 $\mu$ V/m
Alcance dinámico diferencial	80 dB (entre dos estaciones utilizables de una cadena)
Tiempo máximo de autoadquisición	450 s (90% de probabilidad)
Desviación estándar del error de diferencia de tiempo	1 $\mu$ s

Además, el equipo debe poder seguir razones de diferencias de tiempo adecuadas para el tipo de aeronave que se vaya a utilizar (hasta un cambio de 8  $\mu$ s/s a 2 500 nudos). Típicamente se pueden lograr exactitudes de diferencia estática de tiempo de 100 ns y dinámica de 300 ns.

Un sistema de recepción LORAN de a bordo relativamente simple que exhibe al piloto conjuntos de diferencias de tiempo en microsegundos pesa unos 8.5 kg y consume 50 W de potencia. Sin embargo, la salida primaria es una corriente de datos en serie que se envía a una computadora de navegación situada a bordo, para combinarlos con otras salidas de los sensores de a bordo. Suele emplearse la antena de sentido del radiogoniómetro automático (ADF) de la aeronave para recibir señales de LORAN.

**LORAN-D.** El LORAN-D es una versión táctica del LORAN-C que es operada por la fuerza aérea de Estados Unidos. Las estaciones de tierra son transportables por aire, lo que hace posible la recolocación de una cadena en menos de 48 horas. Los pulsos del LORAN-D se envían en grupos de 16, separados 500  $\mu$ s. El ajuste de fases dentro de cada pulso se puede codificar por medio de dispositivos criptográficos para evitar el uso no autorizado de las señales. Los receptores LORAN-D a menudo tienen la capacidad de manejar ambos tipos de señales C y D. Durante el procesamiento de señales del LORAN-D algunas veces se utiliza el procesamiento de correlación de muchos pulsos, a fin de reducir en forma sustancial el ancho de banda de procesamiento efectivo.

**Navegación Doppler.** Los principios se exponen en la sección 51.8. La RTCA ha establecido las normas de funcionamiento<sup>31</sup>. En los primeros sistemas se utilizaban antenas estabilizadas en uno o dos ejes para compensar el balanceo y cabeceo de la aeronave, pero ahora son más comunes los sistemas de antena fija con compensación en el procesamiento de las señales. El radar Doppler genera señales proporcionales a las ve-



**Fig. 52-19.** Unidad de visualización (CDU) CMA-771 OMEGA. Cortesía de Canadian Marconi Co. Avionics Division, Montreal, Quebec. La CDU presenta visualizaciones integrales de datos de navegación y guiaje, conservando la sencillez de operación. Una variante del CDU CMA-771 incluye una capacidad de carga de datos remotos. En la posición de prueba del interruptor de selección de visualización, se pueden monitorizar diversas funciones de autoprueba y también condiciones de señales y estaciones, entradas externas al sistema, estimación de errores de posición y revisiones de mantenimiento mientras el sistema prosigue la navegación.

- 1 INTERRUPTOR ON/OFF (encendido y apagado)
- 2 INTERRUPTOR SELECTOR DEL MODO DE VISUALIZADO: hora media de Greenwich/fecha (GMT/D), aproamiento ángulo de deriva (HDG/DA), trayectoria deseada/trayectoria real (DSRTK/TK), distancia de trayectoria transversal/error de ángulo de trayectoria (XTK/TKE), latitud y longitud de posición (POS), punto de ruta (WPT), rumbo/distancia (BRG/DIS), velocidad respecto a tierra/tiempo estimado al punto de ruta (GS/ETE), prueba de dirección y velocidad del viento (WIND)
- 3 CONTROL DE INTENSIDAD DE ILUMINACIÓN (DIM)
- 4 VISUALIZACIÓN DEL NÚMERO DEL PUNTO DE RUTA (WPT)
- 5 VISUALIZACIÓN DEL PUNTO DE RUTA DESDE/HACIA (FR TO)
- 6 ANUNCIADORES (izquierda o derecha, norte o sur)
- 7 VISUALIZADOR IZQUIERDO
- 8 TECLA DE DEFINICIÓN DEL PUNTO DE RUTA (WPT DEF)
- 9 ANUNCIADORES (izquierda o derecha, este u oeste)
- 10 VISUALIZADOR DERECHO
- 11 ANUNCIADORES: falla del sistema (SYS), modo de operación de navegación a la estima (DR), ambigüedad en la posición (AMB), condición de sincronización Omega (SYN), sistema en modo VLF (VLF)
- 12 TECLADO DE DATOS
- 13 TECLA DE RETROCESO (BK)
- 14 TECLA DE RETENCIÓN EN POSICIÓN (HLD)
- 15 TECLA DE CAMBIO DE RUMBO (LEG CHG)
- 16 INDICADOR DE ENTRADA
- 17 TECLA DE ENTRADA (ENT)
- 18 INTERRUPTOR DE CAMBIO DE RUMBO AUTOMÁTICO/MANUAL (AUTO/MAN). En una variante del CDU se incorpora un interruptor remoto automático/manual

locidades longitudinal y transversal, y también a la velocidad vertical en algunos casos. Para deducir la posición por integración de estas velocidades se utiliza un indicador de rumbo exacto externo. Los errores de velocidad típicos son del orden del 0.3% ( $2\sigma$ ). El error longitudinal típico es del 0.5% de la distancia recorrida en vuelo (95% de probabilidad), y el error en la medición del ángulo de deriva es de  $0.75^\circ$ , excluyendo el error del indicador de rumbo.

La unidad AN/APN 200-V mide aproximadamente  $41 \times 41 \times 13$  cm y pesa 16.5 kg. Cada antena de recepción y transmisión individual es un montaje plano de guías de ondas ranuradas, que producen tres haces con anchos de  $5^\circ$  en el plano de la depresión y  $11^\circ$  en el plano del lado ancho, con un ángulo de depresión de  $67^\circ$ . La señal radiada es de onda continua a una frecuencia aproximada de 13.3 GHz y con la potencia de 100 mW. Las salidas de frecuencia Doppler de los tres canales de recepción son frecuencias de audio que se procesan para producir señales proporcionales a las velocidades longitudinal, transversal y vertical. El consumo total de potencia de esta unidad es de unos 220 W.

En su configuración más simple, los datos Doppler se combinan con entradas de rumbo de la aeronave para facilitar una visualización de la distancia y el rumbo integrados respecto a un punto de partida conocido. En el caso de los sistemas de líneas aéreas y militares, las señales Doppler se envían a una computadora de navegación y se combinan con datos de rumbo, altitud y otros datos aéreos. La visualización se produce entonces en forma de instrucciones de guiaje a izquierda y derecha, y distancia por recorrer hasta un punto de la ruta seleccionado, o coordenadas geográficas en un tipo de visualizado "de mapa en movimiento". La navegación Doppler se utiliza intensamente en helicópteros militares y civiles. Con sistemas optimizados para los intervalos de velocidad de estos vehículos, se puede utilizar para el control del vuelo estacionario.

### 52.2.4 Equipo de detección meteorológica

#### Radar de a bordo

Muchas aeronaves militares están equipadas con radar de mapeo del terreno para su uso como auxiliar de la

navegación. Sin embargo, el radar de a bordo no suele utilizarse para este fin en las aeronaves civiles. Más bien se emplea como auxiliar en la evasión o penetración de malas condiciones climatológicas. La instalación de alguna forma de auxiliar para la evasión de malas condiciones meteorológicas es requerida por la FAA en todas las aeronaves de líneas aéreas y en la mayor parte de las aeronaves comerciales de Estados Unidos.

Los radares meteorológicos de a bordo detectan señales reflejadas por gotas de lluvia y granizo. En regiones de tormentas eléctricas y gran turbulencia, el tamaño y la intensidad de las gotas son mucho mayores que en las nubes estratiformes estables o niebla. Como la respuesta dada por el radar, con base en la dispersión de Rayleigh, varía como la sexta potencia del diámetro de las gotas, la detección de señales intensas reflejadas se correlaciona bien con regiones de turbulencia. Muchos radares operan en el modo "de contorno", en el cual la visualización se presenta en términos del gradiente de la intensidad de señales reflejadas con la distancia. Un gradiente de alto valor absoluto es un buen indicador de turbulencia.

La RTCA ha establecido normas de funcionamiento de radares meteorológicos de aviones<sup>32</sup>. El equipo opera a unos 5 GHz, o más comúnmente en la actualidad a unos 9.35 GHz. El ancho de los pulsos varía de 2 a 5  $\mu$ s. Las potencias pico varían de 3 a 5 kW en el caso de equipo diseñado para aeronaves bimotores livianas y hasta 50 kW o más para unidades de líneas aéreas. En el equipo de líneas aéreas ahora en proceso de desarrollo se utilizan potencias pico mucho menores, pero con longitudes de pulso mayores para conservar la sensibilidad y con procesamiento de correlación para preservar la resolución de distancia. Las antenas varían desde los 30 cm de diámetro hasta unos 80 cm, según el tamaño del avión. Para instalaciones mayores, la antena se estabiliza contra el balanceo y cabeceo de la aeronave. La amplitud de exploración es comúnmente  $\pm 60^\circ$  en sentido horizontal con respecto a la nariz del avión, a un ritmo aproximado de 12 a 20 barridos por minuto. Por lo general, el ángulo de la antena de haz en forma de pincel se puede inclinar por encima o por debajo del horizonte hasta  $\pm 15^\circ$  bajo control manual del piloto, a fin de investigar la extensión vertical de las señales reflejadas por precipitación. A ángulos de depresión bajos, el radar puede emplearse asimismo para mapear el terreno y como auxiliar para la navegación. El equipo típico para las líneas aéreas pesa 40 kg y consume 250 W; un sistema para aeronave pequeña por lo común pesa menos de 10 kg y consume menos de 100 W. Las distancias utilizables máximas varían de 120 hasta 300 nmi para el equipo de líneas aéreas.

Aunque muchos sistemas de radar meteorológico que se siguen empleando producen una exploración en pantalla de CRT por sectores rho-theta, los radares meteorológicos modernos realizan el almacenamiento y conversión del barrido para generar una imagen de exploración tipo X-Y estable. Ahora se está usando el color para representar las diferentes intensidades de señal reflejada.

La tendencia actual consiste en utilizar el CRT del radar meteorológico para combinar datos del clima con otros datos de navegación en una sola visualización, y utilizarlo también para la presentación de caracteres alfanuméricos, tales como listas de verificación y mensajes de emergencia.

#### DetECCIÓN y telemetría de descargas eléctricas atmosféricas

En estudios realizados por la fuerza aérea de Estados Unidos y la NASA se ha demostrado que existe una correlación importante entre condiciones turbulentas asociadas con tormentas eléctricas y la posición e intensidad de descargas eléctricas. Además del efecto visible, estas descargas incluyen muchos fenómenos invisibles que se pueden detectar mediante el uso de receptores de radio de baja frecuencia.

Un dispositivo comercial diseñado para detectar turbulencia midiendo estos efectos es el "Stormscope" (3M-Ryan Corp., Columbus, Ohio). Es un sistema sólo de recepción en el que la dirección de llegada de la energía de descarga se mide mediante el uso de una antena similar a la del ADF. Se monitorizan frecuencias seleccionadas en una banda en torno a los 50 kHz. La medición de la distancia se realiza por medio de una técnica patentada. La intensidad de las señales, tiempos de ascenso y descenso de pulsos, y diferencias entre las componentes de los campos E y H se miden y comparan con conjuntos de modelos de "impresiones dactilares" prealmacenados que representan diversas condiciones de medición. Las coordenadas rho-theta de cada descarga detectada se presentan como un punto en un CRT de almacenamiento, lo que produce un despliegue de  $360^\circ$  en acimut en torno a la aeronave. Las agrupaciones densas de puntos indican una actividad intensa y por tanto de turbulencia probable. La exactitud telemétrica es aproximadamente del 10% con distancias máximas de 250 nmi.

Aunque originalmente se desarrolló como un auxiliar meteorológico para aquellas aeronaves ligeras en las cuales no era practicable la instalación de un radar meteorológico, el Stormscope está siendo cada vez más utilizado en aeronaves de mayor tamaño, hasta el punto de que ha sido certificado por la FAA como un auxiliar en la evasión de ciertas condiciones meteorológicas y aprobado para ciertas operaciones comerciales en Estados Unidos.

#### REFERENCIAS BIBLIOGRÁFICAS

1. *Federal Radio Navigation Plan*, US Department of Defense and US Department of Transportation, DOT-TSC-RSPA-80-16, vols. 1-4.
2. P. C. Sandretto, *Electronic Navigation Engineering*, International Telephone and Telegraph Corp.
3. H. C. Hurley, S. R. Anderson y H. F. Keary, "The CAA VHF Omnidirectional", *Proc. IRE* 39:1506-1520.
4. *Aeronautical Telecommunications*, Anexo 10, Convention of International Civil Aviation. ICAO. Montreal, Canadá, 3.<sup>a</sup> ed.



5. "United States National Aviation Standard for the Very High Frequency Omnidirectional Radio Range (VOR)/Distance Measuring Equipment (DME)/Tactical Air Navigation (TACAN) Systems", FAA Advisory Circular, AC-00-31.
6. A. Alford y A. G. Kandoian, "Ultra High Frequency Loop Antenna", *Electr. Commun.* 18(5).
7. S. R. Anderson, "VHF Omnirange Accuracy Improvements", *IRE Trans. Aero. Nav. Electr. ANE-12*(1):26-35.
8. *Minimum Performance Standards-Airborne VOR Equipment*, DO-153A, RTCA, Washington, DC.
9. S. R. Anderson y R. B. Flint, "The CAA Doppler Omnirange", *Proc. IRE* 47(5, Pt. 1):808-821.
10. W. J. Crone y E. L. Kramar, "Development of the Doppler VOR in Europe", *IEEE Trans. Aero. Nav. Electr. ANE-12*(1):36-40.
11. S. H. Dodington, "Development of 1000 MHz Distance Measuring Equipment", *IEEE Trans. Aero. Electr. Syst. AES-16*(4).
12. M. Kayton y W. R. Fried, eds., *Avionics Navigation Systems*, Wiley, Nueva York.
13. *Minimum Performance Standards-Airborne DME*, DO-151A, RTCA, Washington, DC.
14. P. C. Sandretto, "Development of TACAN at Federal Telecommunication Laboratories", *Electrical Communication* 33(1):4-10.
15. "Standard Tactical Air Navigation (TACAN) Signal", US Military Standard MIL-STD-291B.
16. *Minimum Performance Standards-Airborne ADF Equipment*, DO-142, RTCA, Washington, DC.
17. "The Decca Navigator System and Its Uses", Pub. K20/3, Decca Navigator Co., Ltd., Londres.
18. "Selection Order: U.S. National Standard for the IFF Mark X (SIF) Air Traffic Control Radar Beacon System (ATCRBS) Characteristics", DOT/FAA Order 1010.51.
19. "Design for the National Airspace Utilization System", FAA, Washington, DC.
20. *Minimum Performance Standards-Airborne ATC Transporter Equipment*, DO-150 RTCA, Washington, DC.
21. "Discrete Address Beacon System (DABS) National Standard", FAA Order #6365.1.
22. V. A. Orlando y P. R. Drouilhet, "Discrete Address Beacon System (DABS) Functional Description", MIT Lincoln Labs, FAA-RD-80-41, NTIS AD-A085169.
23. Characteristic 718-2, "Mark 3 Air Traffic Control Transponder (ATCRBS/DABS)", Aeronautical Radio Inc. (ARINC), Annapolis, MD.
24. J. D. Welch y V. A. Orlando, "Active Beacon Collision Avoidance System-Functional Overview", MIT Lincoln Labs, FAA-RD-80-127, NTIS AD-A094177.
25. "Briefing Spurs Carrier Issue of Beacon", *Aviat. Week Space Technol.*
26. R. L. Eisenberg, "JTIDS System Overview", en C. T. Leondes, ed., *Principles of Precision Position Determination Systems*, AGARDograph n.º 245.
27. G. I. Palattucci, "JTIDS Signal Structure", en C. T. Leondes, ed., *Principles of Precision Position Determination Systems*, AGARDograph n.º 245.
28. "Long Range Navigation Market Thriving", *Aviat. Week Space Technol.*
29. *Minimum Performance Standards-Airborne Omega Receiving Equipment*, DO-164A, RTCA, Washington, DC.
30. *Minimum Performance Standards-Airborne LORAN Equipment*, DO-159, RTCA, Washington, DC.
31. *Minimum Performance Standards-Airborne Doppler Radar Navigation Equipment*, DO-158, RTCA, Washington, DC.
32. *Minimum Performance Standards-Airborne Weather and Ground Mapping Pulsed Radars*, DO-134, RTCA, Washington, DC.

## 52.3 AUXILIARES DE LA NAVEGACIÓN MARÍTIMA

Thomas J. Cutler

### 52.3.1 Introducción

La ciencia de determinar la ubicación de una nave en el mar (navegación) tiene muchas facetas, y va de lo muy simple a lo extremadamente complejo. No obstante, el denominador común de casi todos los sistemas de navegación es la "línea de posición" (LOP). El simple pilotaje mientras se tiene visión de la tierra se realiza observando un objeto conocido representado en una carta de navegación y una brújula a fin de determinar un rumbo visual en línea recta. Ese rumbo, cuando se transfiere a una carta, es una LOP. La ciencia más bien compleja de la navegación astronómica (celeste o sideral) se vale del ángulo de altitud de cuerpos astronómicos (que se obtiene por medio del sextante) para determinar un "círculo de igual altitud", que es simplemente otra forma de LOP.

La LOP por sí misma no revela la posición al navegante, sino más bien da un *indicio* de la ubicación. El navegante sabe que se encuentra en alguna parte de esa LOP, pero no puede determinar exactamente dónde con un grado de precisión cualquiera si no cuenta con más información. No obstante, si el navegante puede obtener dos o más LOP, sabrá dónde se cortan en su ubicación, que se llama marcación, punto de posición, posición relativa o punto fijo.

La navegación electrónica no es diferente. A pesar de los grados de complejidad y avance tecnológico, casi todos los sistemas tienen sólo un fin básico: ofrecer al navegante una o más LOP. La forma en que se logra esto varía considerablemente en un vasto espectro de sistemas y diseños de equipo. Algunos de los sistemas electrónicos que utilizan los marinos para ayudarse en la navegación, específicamente sistemas de radar, sonar y satélite, se analizan en otras secciones de este manual porque sus aplicaciones cubren mucho más allá de la ciencia de la navegación. Sin embargo, aquí se

Tabla 52-4. Comparación de sistemas de navegación marítima

	RDF	LORAN-A	LORAN-C	OMEGA	DECCA	CONSOL
Frecuencia (kHz)	285 a 325	1850 a 1950	90 a 110	10.2 a 13.6	70 a 130	190 a 370
Área cubierta	La mayor parte de las regiones costeras hasta 175 nmi	Japón	La mayor parte de las regiones costeras hasta 2 000 nmi	Virtualmente todo el mundo	Este de Canadá; Europa Occidental; Japón; Golfo Pérsico; partes de India, África, Australia	URSS; Europa (de Noruega a España); Estados Unidos (Nantucket, San Francisco, Miami)
Tipo	Medición de rumbo radial	Diferencia de tiempo hiperbólica	Diferencia de tiempo hiperbólica	Fase hiperbólica	Fase hiperbólica	Hiperbólico/radial
Alcance máximo (nmi)	175	800	1500	Virtualmente ilimitado	240	1400 (mín. de 25 a 30)
Mayor exactitud (nmi)	Variable	0.5	0.25	1	0.1	1
Comentarios	El menos costoso	En desuso; sólo en Japón está indeterminada la fecha en que se discontinuará	Reconocido como el principal sistema electrónico de navegación de Estados Unidos	El más costoso; el más nuevo	De propiedad privada	"CONSOL" en Europa; "CONSOLAN" en Estados Unidos; la URSS tiene un sistema comparable, pero distinto

<sup>a</sup> RDF, radiogoniometría; LORAN, navegación de larga distancia; nmi, millas náuticas (1 nmi = 2 000 yd = 1 828.8 m).

Tabla 52-5. Uso del espectro de frecuencia para la navegación marítima

Banda	Frecuencia	Uso para la navegación marítima <sup>a</sup>	Otros usos
EHF	30 a 300 GHz		Comunicación por microondas
SHF	3 a 30 GHz		
UHF	300 a 3000 MHz		Radar
VHF	30 a 300 MHz		NAVSTAR
HF	3 a 30 MHz		TV
MF	300 a 3000 kHz	LORAN-A	Navegación por satélite
LF	30 a 300 kHz	RDF CONSOL LORAN-C	Radio AM
VLF	3 a 30 kHz	DECCA OMEGA	

<sup>a</sup> LORAN, navegación de larga distancia; RDF, radiogoniometría.

analiza una gran cantidad de sistemas utilizados exclusivamente para la navegación marítima.

Debe ponerse de relieve el hecho de que ningún navegante que se precie de serlo confía en una sola fuente de información para determinar puntos de posición. Todos los métodos de obtención de la(s) LOP tienen limitaciones. Los sistemas electrónicos están sujetos a fallas o fluctuaciones de energía, condiciones atmosféricas y errores del operador. La navegación astronómica puede verse restringida por nubes que oculten al observador los cuerpos celestes, y el pilotaje costanero es imposible si hay niebla. Además, la vital importancia de lo que él hace y las consecuencias de las fallas indican al navegante que nunca debe estar satisfecho sino hasta que haya agotado todas las fuentes de información disponibles para convencerse de que su embarcación no está en peligro.

#### Sistemas electrónicos

Los sistemas electrónicos exclusivos para la navegación marítima son la radiogoniometría (RDF), la navegación de larga distancia (LORAN), OMEGA, DECCA y CONSOL. En la tabla 52-4 se comparan estos sistemas. Los alcances y exactitudes que se indican son valores ideales teóricos, y no necesariamente se observarán en la realidad en condiciones subóptimas; sólo se incluyen para facilitar la comparación de los sistemas. Los sistemas "hiperbólico" y "radial" se explican en una subsección posterior.

Otro sistema de navegación, el sistema de navegación por inercia (o inercial) de barcos (SINS), que también se analiza en una subsección posterior, no está incluido en la tabla 52-4 porque su carácter único no permite hacer una comparación significativa con los otros sistemas.

**Frecuencias.** Todos los sistemas que se analizarán (salvo el SINS) se basan en la transmisión de señales de radio por la atmósfera terrestre, y por tanto ocupan porciones del espectro de frecuencias de radio. En la tabla 52-5 se ilustra la relación del uso del espectro para la navegación marina con el uso que hacen de él otros sistemas electrónicos.

Se puede apreciar que la navegación electrónica está confinada a la porción inferior del espectro RF (es decir, MF y más abajo), salvo en la navegación por satélite.

Las muy altas frecuencias (VHF) y mayores están limitadas en alcance a la línea de visión (trayectoria óptica). Las frecuencias medias y altas (MF y HF) son vulnerables a las pérdidas elevadas de propagación cuando viajan sobre tierra firme, y la predictibilidad de la propagación puede ser errática si en la trayectoria de transmisión se encuentran con estructuras conductoras creadas por el hombre, cuyo tamaño sea una fracción importante de la longitud de onda que se usa. Las transmisiones de baja frecuencia (LF) son mucho más estables y por tanto predecibles, y debido a su longitud de onda grande son adecuadas para hacer comparaciones exactas de tiempo y fase. Sin embargo, a mayores distancias, las ondas celestes (reflejadas) comienzan a volver a la superficie de la Tierra y causan interferencia con la onda terrestre, por lo que se necesita algún medio de discernimiento entre las dos (fig. 52-20). Las muy bajas frecuencias (VLF) viajan distancias extremadamente largas en la guía natural de ondas que se produce entre la superficie de la Tierra y las capas ionosféricas de la atmósfera. El alcance es de gran ventaja, pero la necesidad de tener conocimiento en tiempo real de las condiciones ionosféricas complica el uso de estas frecuencias.

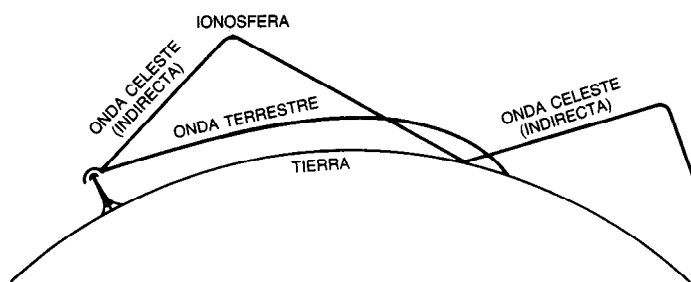


Fig. 52-20. Propagación de ondas terrestres y celestes (reflejadas).

En la tabla 52-5 se puede apreciar que la porción LF del espectro de frecuencia de radio ha sido la más popular en los sistemas de navegación. El LORAN-A comenzó en la región MF, pero el LORAN-C, su sucesor mejorado, optó por la porción LF, más estable. Sólo OMEGA ha figurado en la región VLF; el deseo de crear un sistema de navegación mundial con un número mínimo de puntos de transmisión requeridos fue suficiente para superar las desventajas del aumento de los costos y la complejidad del diseño.

**Sistemas radiales.** En los primeros intentos que se hicieron por incorporar la electrónica a la navegación se emplearon antenas radiogoniométricas situadas en la costa que podían tomar rumbos direccionales de barcos que transmitían sus frecuencias de radio en el mar. Después, la estación de la costa podía enviar esta información al barco, con lo cual se le daba una LOP. Dos o más de estas estaciones costeras proporcionaban un punto fijo de posición.

Una evolución posterior invirtió los papeles al colocar la antena radiogoniométrica a bordo de la embarcación y convertir las estaciones de la costa en sitios de transmisión. El resultado obtenido fue el mismo, pero hizo posible que el barco operara sólo en modo de recepción o pasivo, y no requería que se establecieran comunicaciones. Ésta es la radiogoniometría (RDF), que se analiza con mayor detalle en una subsección posterior.

El principio de obtención de un rumbo en línea recta desde un solo objeto convierte a éste en un sistema radial, ya que el rumbo se obtiene desde un transmisor que emite señales en todas direcciones.

**Sistemas hiperbólicos.** Los sistemas hiperbólicos se producen colocando dos estaciones de transmisión separadas por una distancia considerable y haciendo que éstas transmitan una señal reconocible. Una línea teórica entre las dos estaciones se denomina línea base, y la que continúa más allá de las estaciones es la extensión de la línea base (fig. 52-21a). Si una embarcación está equipada para recibir las señales y discernir alguna diferencia en el tiempo en que se recibe cada señal, se puede obtener una indicación de la posición del barco (es decir, una LOP). Por ejemplo, si el barco recibe ambas señales transmitidas al mismo tiempo (o sea, sin diferencia de tiempo), entonces la embarcación debe estar en alguna parte a lo largo de la bisectriz perpen-

dicular de la línea base (indicada por "Cero" en la fig. 52-21a). La falta de una diferencia de tiempo indica al navegante del barco que está situado en algún punto equidistante de las dos estaciones transmisoras. Tan pronto como la embarcación ya no se encuentre exactamente a la misma distancia de cada una de las estaciones, la señal de una estación llegará antes que la otra (habrá una diferencia en tiempo discernible en la recepción de las señales). Una vez más, esto indica al navegante que se encuentra en alguna parte a lo largo de una línea definible. Pero como se trata de diferencias en el sentido matemático, la línea resultante no es recta, sino hiperbólica. Cada combinación de diferencias de tiempo (medidas en microsegundos) da por resultado una hipérbola única (fig. 52-21b). Mediante la utilización de una carta (mapa) de navegación con las hipérbolas adecuadas superpuestas, el navegante puede medir la diferencia de tiempo entre la recepción de las señales de cada estación a fin de identificar la hipérbola adecuada. Este proceso produce una sola LOP.

Para obtener un punto de posición, se debe emplear otro conjunto de estaciones con su propia hipérbola única (fig. 52-21c). Allí donde se cortan las hipérbolas se encuentra el punto de posición.

Como la sincronización entre las dos estaciones es tan crítica, en los diseños originales se empleaba un concepto de "amo" y "esclavo", donde la estación maestra o amo no sólo transmitía su señal, sino que inducía la activación del esclavo para asegurar simultaneidad; de aquí las designaciones "M" y "S" de las estaciones. El concepto de esclavo ha evolucionado a "secundario" con el advenimiento de los relojes modernos extremadamente exactos, que hacen innecesaria una activación dependiente.

Las ventajas del sistema hiperbólico son significativas. No se requiere una antena direccional; una simple antena de alambre es todo lo que se necesita. Aunque resulta altamente complejo, el equipo que se necesita a bordo es relativamente compacto y no requiere capacidad de transmisión como el aparato del radar. El alcance y exactitud de un sistema hiperbólico resultan muy superiores a los de cualquier sistema acimutal simple.

Aunque no es muy evidente para el ojo humano, las líneas que figuran en la carta de navegación no son hipérbolas reales debido a las distorsiones ocasionadas por la forma esferoidal de la Tierra.

### 52.3.2 Auxiliares electrónicos de la navegación

#### Radiogoniometría (RDF)

Como ya se dijo, la radiogoniometría (RDF) fue el primero de los métodos de navegación electrónica y se basa en el principio acimutal o radial. Las LOP que se obtienen son rumbos de simples líneas rectas que se trazan fácilmente en casi cualquier carta de navegación. El equipo RDF es el menos costoso de todos los componentes de navegación electrónicos, y consta únicamente de un receptor y una antena radiogoniométrica. Todas las embarcaciones de más de 1 600 toneladas de peso bruto y en funciones de comercio internacional están obligadas, por el convenio de seguridad de la vida en el mar (SOLAS) de 1960, a contar con un radiogoniómetro.

**Antenas.** La antena radiogoniométrica en su forma más simple es un cuadro que modifica la intensidad de una señal recibida en función de su ángulo con respecto a las ondas de radio incidentes. Cuando el cuadro es paralelo a las ondas, la señal es más intensa, y cuando es perpendicular, la señal es más débil o completamente imposible de identificar. Este punto de mínima señal se conoce como punto cero o de anulación, y es el que se utiliza para determinar el rumbo hacia la estación transmisora. Existe cierta ambigüedad inherente a un sistema de este tipo, porque, al girar 360°, la antena llega dos veces al punto cero, cada uno separado 180°. Uno es la orientación real a la estación y el otro es el recíproco. La confusión potencial se puede eliminar uniéndolo a una antena extra "de sentido" que diferencia entre las orientaciones real y recíproca.

El término "ADF" (radiogoniometría automática) se aplica a equipos que además tienen las características de una antena que se hace girar eléctrica o mecánicamente y una visualización que presenta una lectura continua del rumbo (corregida automáticamente en cuanto a ambigüedad), que se analizó antes. En el equipo RDF naval moderno se utiliza una antena continuamente giratoria impulsada por un motor, la cual monitoriza la intensidad de la señal comparativa y presenta la información de rumbo resultante sobre un indicador de pantalla de tubo de rayos catódicos (CRT). Los sistemas RDF y ADF se utilizan asimismo en los sistemas de navegación para la aviación (sec. 52.2).

**Estaciones.** Las estaciones transmisoras en tierra se denominan radiofaros marítimos, y a menudo se combinan con auxiliares de la navegación visuales ordinarios, tales como boyas o faros. Normalmente transmiten una señal de onda continua de baja potencia modulada por clave Morse. A los radiofaros marítimos se asignan frecuencias específicas en el intervalo de 285 a 325 kHz. Idealmente, cada estación opera en una frecuencia discreta, pero cuando esto no es posible las estaciones se colocan con la suficiente separación geográfica para evitar interferencias. En algunos casos esta separación tampoco es posible y en su lugar se utiliza un sistema de tiempo compartido.

Además de las estaciones de radiofaros creadas específicamente para uso de la navegación marítima,

existen otras fuentes de información de RDF disponibles para el navegante. Se pueden emplear estaciones de radio comerciales siempre y cuando se muestre la ubicación de la antena transmisora en la carta de navegación que se utilice.

Los radiofaros de aeronaves también pueden proporcionar información de rumbo confiable; aunque estos radiofaros cambian gradualmente a frecuencias mayores que escapan al alcance de la mayor parte de las unidades RDF marinas, un número importante de ellas se sigue recibiendo.

En algunos países todavía existen estaciones que aplican el viejo método de RDF, donde el barco hace la transmisión y la estación de tierra interpreta la dirección y la envía a la nave. Estas estaciones se denominan estaciones de radiobrájula (radiocompás o radiogoniómetro).

**Publicaciones.** Es posible encontrar datos sobre ubicación, frecuencia, distancia y característica de identificación (es decir, el identificador en clave Morse de radiofaros marítimos ubicados en Estados Unidos) en la publicación de la guardia costera de ese país, *Light List*, la cual, a pesar de su título más bien restrictivo, contiene información sobre radioayudas de navegación, señales de niebla, boyas, buques faro y balizas de día, así como sobre luces. Se divide en cinco volúmenes separados según la región y se actualiza continuamente a través del sistema de "Notice to Mariners" (aviso a los marinos).

La información de RDF para aguas extranjeras está contenida en la publicación 117 de Defense Mapping Agency, *Radio Navigational Aids* (en un principio HO117, cuando la U.S. Navy Hydrographic Office tenía la responsabilidad de su publicación). Existen dos volúmenes aparte, 117A y 117B, que se aplican a las áreas del mar Mediterráneo-Atlántico y del océano Pacífico-Índico, respectivamente.

**Exactitud.** Una señal de radio que viaja sobre tierra está sujeta a algún grado de distorsión en el rumbo. Por tanto, cuanto más al interior de tierra esté situada una estación, mayor será la pérdida de exactitud. El marino deberá tener conocimiento de esta anomalía cuando seleccione estaciones, porque muchas de las estaciones de transmisión comerciales y radiofaros de aeronaves están localizadas muy al interior de tierra, lo que significa que existe un potencial de error importante.

Las componentes estructurales de la embarcación misma pueden ocasionar interferencia que afectará negativamente la exactitud. Las señales que llegan pueden ser recibidas y retransmitidas por objetos metálicos como la jarcia muerta (firme) de la embarcación (soportes de alambre del mástil, etc.). La existencia de estas señales falsas en la proximidad de la antena afecta negativamente la capacidad del receptor de determinar la dirección.

Un efecto de polarización es evidente durante las horas de la noche, en particular cerca de las horas de salida y puesta de Sol, lo que genera cierta inexactitud que no se encuentra a la luz del día.

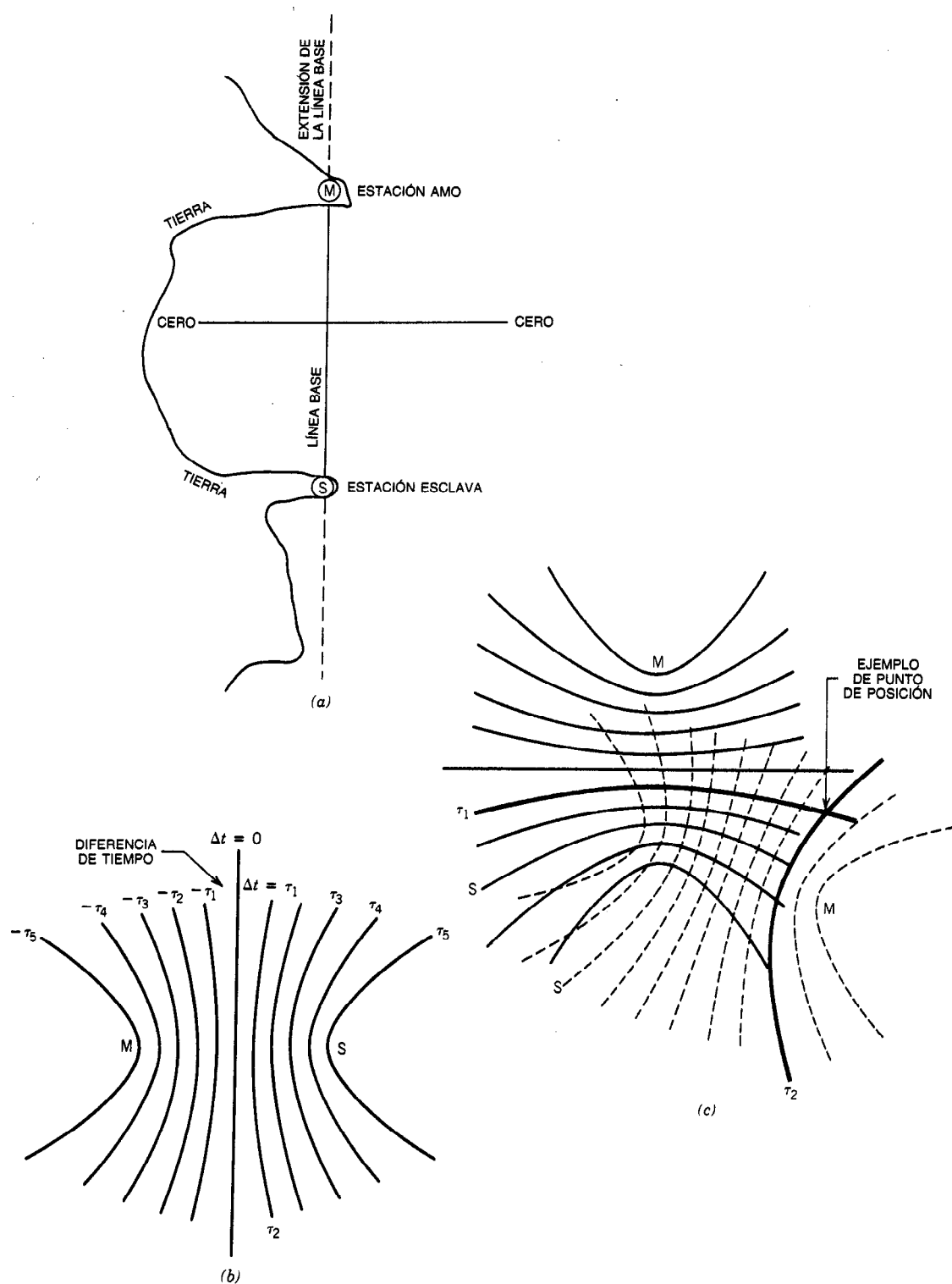


Fig. 52-21. Sistema de navegación hiperbólico: a) geometría básica; b) diagrama hiperbólico; c) sistemas hiperbólicos que se cortan.

Otros factores que pueden afectar la exactitud de los rumbos del RDF son la intensidad de la señal (cuanto más débil, menos exacto es el rumbo) y la capacidad del marino de utilizar el equipo (la práctica y la experiencia mejoran considerablemente los resultados).

**Radiofaro omnidireccional.** Diseñado como un sistema de navegación de aeronaves, el radiofaro omnidireccional (abreviado omni) es utilizado por algunos marinos que tienen la posibilidad financiera de poseer el equipo receptor considerablemente más costoso que se requiere. El sistema opera en la escala VHF (108 a 118 MHz) y es compacto, ligero y simple de usar. Existe un número importante de estaciones ubicadas lo suficientemente próximas a la línea costera para permitir a las embarcaciones su uso, pero la cobertura no es de ningún modo continua en las vías acuáticas de Estados Unidos.

Este sistema puede emplearse para obtener rumbos (y por tanto LOP), pero es más efectivo cuando se emplea como sistema de aproximación por radio; es decir, el marino elige una estación localizada en o cerca de su destino y después utiliza el receptor omnidireccional para mantener el rumbo hacia el transmisor. Debido a la característica de cercanía de trayectoria óptica de las señales VHF, el alcance está más bien limitado para las embarcaciones de superficie. En la sección 52.2 se expone con más detalle el sistema de radiofaro omnidireccional.

**LORAN.** La palabra "LORAN" es el acrónimo de *long range navigation*, "navegación (o radionavegación) de larga distancia". Es uno de los sistemas de navegación hiperbólica y tiene aplicaciones marítimas y aeronáuticas; también se realiza investigación actualmente para su empleo en tierra (es posible su uso en programación del tránsito, monitorización de vehículos de emergencia, ubicación de trenes, etc.).

**Desarrollo.** El sistema LORAN fue creado en el Massachusetts Institute of Technology (MIT) durante la Segunda Guerra Mundial. La versión original se llamaba LORAN-A y se sigue utilizando en forma limitada hasta la fecha. Una versión posterior, el LORAN-B, era del tipo de pulso y ciclo, pero nunca tuvo éxito debido a las tolerancias de propagación insatisfactorias. LORAN-C, una mejora de diseño del LORAN-A original, es el principal sistema LORAN en uso en el presente y ha sido designado por el gobierno de Estados Unidos como el principal sistema auxiliar civil para la navegación en la "zona de confluencia costera" [CCZ, que se define como 50 millas (92.6 km) tierra adentro o fuera, hasta el borde de la plataforma continental, el que sea mayor] hasta el año 2000.

Un sistema de estaciones transportables creado por la fuerza aérea de Estados Unidos, que se caracteriza por un consumo de potencia relativamente bajo y su resistencia a la interferencia electrónica (deliberada o no), fue designado LORAN-D.

La necesidad de ampliar el alcance de LORAN-A estimuló la investigación que por último condujo al desarrollo del LORAN-C. La disminución de la frecuencia para incrementar la distancia y estabilizar las características de propagación requería una mayor salida de potencia en los sitios de transmisión, y los mayores alcances resultantes crearon nuevos problemas para distinguir entre las ondas terrestres y las celestes (reflejadas). El LORAN-C surgió con el empleo de métodos de comparación de fase así como de medición de las diferencias de tiempo.

Las primeras estaciones de LORAN-C comenzaron a operar en 1957, y el LORAN-A comenzó a desaparecer en 1977. El LORAN-A se programó para que desapareciera por completo a inicios de la década de 1980; la única excepción parece ser Japón, donde la terminación programada se cita como "indeterminada" en la publicación 117B de Defense Mapping Agency (*Radio Navigation Aids*).

En la tabla 52-4 las mejoras en alcance y exactitud son evidentes. Los alcances que se citan en la tabla pueden analizarse en forma más significativa, como refleja la tabla de la parte inferior (1 nmi = 2 000 yd = 1828.8 m).

Dado que es un sistema hiperbólico, el LORAN requiere una configuración de dos estaciones para cada LOP que se genere. La distancia posible entre estaciones creció de un máximo de 500 nmi con LORAN-A a 1 500 nmi con LORAN-C. Este incremento en la distancia de la línea base da mayor flexibilidad en la selección del sitio de la estación y también ofrece una cobertura más vasta por medio de las hipérbolas resultantes.

**Estaciones cadenas.** Aunque se requiere un mínimo de dos estaciones, la disposición normal es una cadena que comprende una estación principal y dos, tres o cuatro estaciones secundarias (la principal se designa M y las secundarias se designan W, X, Y, Z). Las configuraciones son determinadas por la cobertura deseada y la disponibilidad de sitios de ubicación de estaciones, pero con el objeto de optimizar la eficiencia se intenta su adaptación a los patrones básicos que se ilustran en la figura 52-22 siempre que sea posible.

Actualmente hay 13 cadenas de LORAN-C en todo el mundo, que constan de 41 estaciones. La Unión Soviética posee un sistema similar que consta de ocho estaciones.

Dentro de cada cadena existen una o dos estaciones de monitorización, que se destinan a vigilar y contro-

	Alcance de la onda terrestre (nmi)	Alcance de la onda celeste (reflejadas) de un salto (nmi)	Alcance de la onda celeste de dos saltos (nmi)
LORAN-A	450-800	1 400	— —
LORAN-C	1 200	2 300	4 000

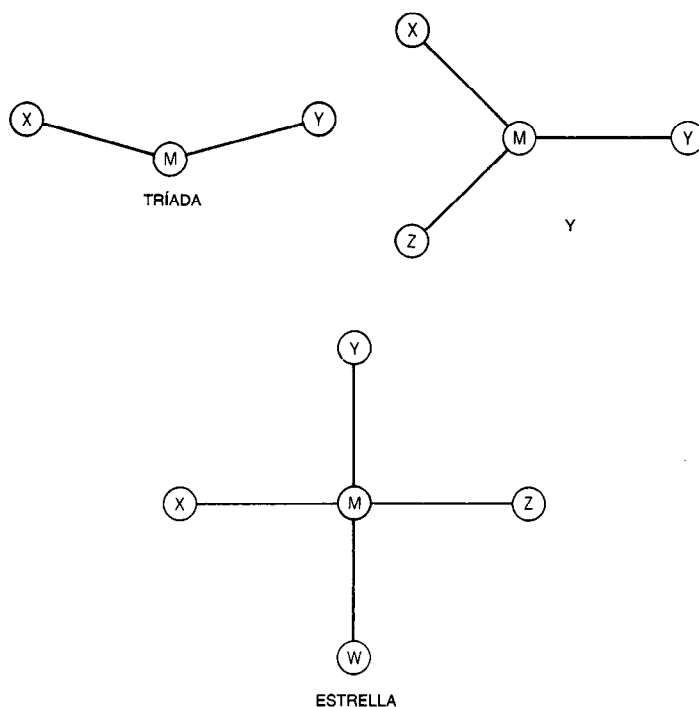


Fig. 52-22. Patrones de disposición de estaciones LORAN. M = estación amo o primaria; X, Y, Z = estaciones secundarias.

lar la operación de la cadena. Este hecho garantiza el formato y la continuidad adecuados de las señales.

La secuencia de transmisión comienza con un pulso de la estación principal, seguido de un pulso de cada una de las secundarias en secuencia, y termina con el inicio de otro pulso de la estación principal. El periodo que transcurre de un pulso principal a otro se denomina intervalo de repetición de grupo (GRI) y es único para cada cadena. Una cadena se identifica por su GRI en microsegundos menos el cero final; por ejemplo, una cadena con GRI de 49 900  $\mu$ s tendría un identificador de 4 990, y las líneas hiperbólicas de la carta de navegación tendrían inscrito este identificador, así como un designador de estación (W, X, Y o Z) y un indicador de diferencia de tiempo (también en microsegundos). Con la información obtenida del receptor y la información en la línea correspondiente de la carta, el marino puede seleccionar la línea (LOP) adecuada entre las representadas, o realizar una interpolación según sea necesario para afinar su posición.

La energía radiada de las diversas estaciones varía de 165 a 1 800 kW. Es interesante observar que la combinación de alta potencia y baja frecuencia permite que las señales de ondas terrestres penetren en realidad bajo el mar hasta cierta profundidad, lo que hace posible que los submarinos las reciban mientras están sumergidos.

**Señales.** Las señales de LORAN son moduladas por pulsos, y se transmiten en grupos de pulsos múltiples

de potencia comparativamente baja. El uso de emisiones pulsadas de este tipo intensifica la capacidad del receptor de discriminar entre las ondas terrestres y las celestes o reflejadas, y también suministra mayor potencia de señal para que el receptor la procese sin tener que incrementar en forma sustancial la salida de potencia en los sitios de transmisión.

Las señales que llegan son comparadas por el receptor en cuanto a diferencia de tiempo, como el principal medio para obtener información LOP, pero una característica extra del LORAN-C incluye una comparación de fase como afinación adicional de la exactitud.

**Errores potenciales.** La antena receptora a bordo de la embarcación debe conectarse a tierra en forma adecuada, y se debe tener precaución al seleccionar un sitio que sea óptimamente elevado en la nave y lo más lejos posible de objetos metálicos.

Los motores o equipos auxiliares pueden generar interferencia, y los aparatos de televisión instalados demasiado próximos a un receptor LORAN pueden producir señales de menor calidad cada vez que reducen la exactitud en forma sustancial o incluso inutilizan por completo el equipo.

Para garantizar una mínima distorsión de la forma de la señal que llega, los receptores LORAN-C tienen un ancho de banda relativamente amplio (normalmente 20 a 30 kHz). Sin embargo, esta característica hace que el equipo sea vulnerable a ruido extraño. Para contrarrestar este inconveniente, los fabricantes inclu-



yen a menudo filtros de muesca en sus equipos. Estos filtros pueden ser *internos* (ajustados por el fabricante) o *sintonizables* (ajustados por el operador).

A pesar de su precisión, los intervalos de sincronización entre el amo y el esclavo pueden ser erróneos. Este "error de sincronización" normalmente se mantiene dentro de  $\pm 0.2 \mu s$  (lo que corresponde a un error de  $\pm 60$  m en distancia) y se considera tolerable. Sin embargo, si el error llega a crecer más de  $0.2 \mu s$ , un sistema de advertencia automático se dispara dentro del transmisor y altera la señal de salida lo suficiente para prevenir a los usuarios y para alertar a las estaciones de monitorización.

**Variaciones del equipo.** Los tipos de receptores varían en complejidad desde un acoplamiento manual de señales en un tubo de rayos catódicos hasta una conversión por computadora de las diferencias de tiempo en una lectura de latitud y longitud. Se dispone de lecturas remotas para quienes deseen tener acceso a la información de navegación en más de un sitio a bordo de su nave, y se han creado "graficadoras de ruta" que registran la ruta de la nave directamente en una carta.

**OMEGA.** OMEGA es la culminación más reciente de la investigación en cuanto al uso del intervalo de muy baja frecuencia (VLF) para producir un sistema de alcance extremadamente grande que virtualmente ofrezca cobertura en todo el mundo. La formación de una guía de ondas natural entre la superficie de la Tierra y la ionosfera a la frecuencia OMEGA elegida de 10 a 14 kHz permite la propagación de la señal a distancias extremadamente grandes. Estas muy bajas frecuencias son asimismo muy estables, y los defasamientos son relativamente predecibles. Esta última característica es de particular importancia porque OMEGA no depende de la medición de una diferencia de tiempo, como el LORAN, sino que es un sistema de comparación de fases. El receptor mide la diferencia en fase entre señales que se transmiten continuamente y al mismo tiempo desde dos o más estaciones para obtener información de posición.

**Estaciones.** A diferencia de lo que ocurre en el LORAN, no existe una relación de amo a esclavo (o secundario). Dos estaciones cualesquiera se pueden combinar para obtener una LOP.

El diseño de OMEGA incluye ocho estaciones ubicadas en Dakota del Norte, Hawai, Liberia, Noruega, Japón, Argentina, Madagascar y Australia. En Trinidad se localizaba una estación temporal mientras se terminaba la de Australia.

La longitud típica de la línea base (la distancia entre estaciones) es de 5 000 a 6 000 millas. Esta línea base extremadamente larga da por resultado un patrón hiperbólico con problemas de divergencia mucho menores a medida que la distancia aumenta desde la línea base.

La potencia radiada normal es 10 000 W (10 kW) en cada estación. Una vez más, la combinación de alta potencia y baja frecuencia hace posible la penetración de señales bajo la superficie del mar, de manera que

OMEGA está disponible para submarinos sumergidos y para embarcaciones de superficie.

Debido a la longitud de onda extremadamente larga de la señal VLF, los tramos de antena de estas estaciones miden aproximadamente 3 500 m de una torre a otra y pesan alrededor de 17 toneladas. De hecho, la antena de Noruega cubre de un lado a otro un fiordo.

**Frecuencias.** Las estaciones OMEGA transmiten en 10.2, 11.05, 11.22 y 13.6 kHz. Cada una de las ocho estaciones transmite en cada una de las cuatro frecuencias, de tal manera que sólo hay cuatro estaciones transmitiendo en un instante dado y cada una en una frecuencia distinta.

**Receptor.** El receptor OMEGA debe ser capaz de discernir la señal de OMEGA en medio de ruido atmosférico normal y de determinar con exactitud la fase de la señal recibida.

Las antenas receptoras a bordo de barcos normalmente son sintonizables, del tipo de látigo vertical, y miden aproximadamente 3 m de altura. Existen configuraciones de antena especiales para las aeronaves.

**Pasillos.** Como las transmisiones viajan en ondas sinusoidales repetitivas, la comparación de fase produce una información de posición que sólo es útil dentro de media longitud de onda (porque el receptor establece la posición de fase respecto al contorno cero más próximo, lo que ocurre dos veces cada ciclo). Estos corredores de media longitud de onda se representan en las cartas de navegación e identifican como pasillos (*lanes*). Por tanto, es imperativo que el marino sepa en un momento dado cualquiera en qué pasillo se encuentra. Los equipos OMEGA incluyen normalmente una función de conteo de pasillos que registra un cambio cada vez que la nave sale de un pasillo y entra en otro. Por supuesto, el marino prudente también llevará la ruta manualmente trazando la trayectoria estimada sobre la base de los rumbos y velocidades que ha venido utilizando la embarcación durante el recorrido. Una técnica conocida como localización de pasillo amplio, que básicamente crea un pasillo artificial mayor para compararlo con uno normal, permite la resolución de la ambigüedad en la mayor parte de los casos y permite al marino determinar qué pasillo está ocupando en ese momento la embarcación.

La sensibilidad del receptor promedio hace posible una exactitud aproximada de 1/100 de pasillo, que recibe el nombre de centicalle o centilane. Esto, desde luego, produce 100 LOP hiperbólicas por pasillo. La media longitud de onda adecuada para la frecuencia de OMEGA produce un ancho promedio de pasillo de 8 nmi (aproximadamente 15 km).

**OMEGA diferencial.** Una exactitud de una milla es aceptable para el marino en los vastos confines del mar abierto, pero deja de serlo al acercarse a zonas costeras. Para remediar este problema se ideó un sistema conocido como OMEGA diferencial.

Como las condiciones de propagación de las ondas de radio son la fuente principal del error de OMEGA,

Tabla 52-6. Uso de frecuencia de DECCA

Señales por comparar	Armónicos	Multiplicadores	Frecuencia de comparación resultante
Amo	$6f^a$	4	
Esclava roja	$8f$	3	$24f$
Amo	$6f$	3	
Esclava verde	$9f$	2	$18f$
Amo	$6f$	5	
Esclava morada	$5f$	6	$30f$

<sup>a</sup>  $f$ , frecuencia fundamental (siempre entre 14.00 y 14.33 kHz).

y debido a que estas condiciones de propagación son relativamente constantes en un área local dada, una estación monitor, capaz de medir errores y determinar correcciones compensatorias, transmite estas correcciones a usuarios del área. La exactitud resultante está dentro de 0.25 nmi (una mejora notable).

**DECCA.** El sistema de navegación DECCA es el único de propiedad privada. La Racal-Decca Company financia la operación del sistema a través de la venta y arrendamiento del equipo de recepción. El DECCA es utilizado por aeronaves y también por embarcaciones de superficie en el mar.

Aunque se inventó en 1937, el DECCA recibió el primer reconocimiento importante en 1944, cuando se utilizó para guiar dragaminas y lanchones de desembarco en las playas de Normandía durante la invasión del día D. La primera estación comercial se construyó al sureste de Inglaterra en 1946.

**Características.** El DECCA emplea una señal de onda continua (CW) no modulada en el intervalo de frecuencia de 70 a 130 kHz. A diferencia del LORAN y OMEGA, que se consideran sistemas de mar abierto, DECCA es en esencia un sistema auxiliar de navegación costera con un alcance nominal de 240 nmi (444 km). Este alcance limitado es compensado en parte por el alto grado de exactitud. DECCA es similar a OMEGA en que utiliza la comparación de fase en vez de la diferencia de tiempo para establecer la posición.

Otra ventaja importante es la facilidad de operación de DECCA. No se requiere ninguna habilidad especial para leer la información de los cuadrantes (llamados decímetros) que están al frente del receptor, y las cartas DECCA tienen retículas que se codifican en colores coincidentes con el color del decímetro que se utilice.

**Estaciones.** Una "cadena" de estaciones normalmente consta de la estación maestra o amo (o primaria), rodeada de tres esclavas (o secundarias) dispuestas en un patrón similar al patrón "Y" de LORAN (fig. 55-22) con aproximadamente 120° de espaciamiento. La longitud de la línea base se encuentra por lo general entre 60 y 120 nmi (111 a 222 km). Las cadenas vecinas a menudo comparten sitios de esclavas comunes.

Las esclavas se codifican en color para coincidir con los colores de los decímetros del receptor (medidores de fase). Los colores que se utilizan (rojo, verde y

morado) son estándares en todo el sistema. Las hipérbolas resultantes del amo y la esclava de color rojo se imprimen *en rojo* en la carta náutica (mapa) de DECCA. Las otras se colorean también en la forma correspondiente. Cada cadena opera en una frecuencia fundamental única entre 14 y 14.33 kHz. Exactamente, la frecuencia que se utiliza es la que identifica o determina la cadena específica.

**Receptor.** El receptor DECCA consta esencialmente de cuatro receptores independientes alojados dentro de una unidad. Los cuatro se fijan en la frecuencia adecuada seleccionando simplemente la "frecuencia fundamental" de la cadena específica. La frecuencia fundamental ( $f$ ) no se radia en realidad, sino que se utiliza una serie de armónicos, como se ilustra en la tabla 52.6. Dentro del receptor las señales que entran se multiplican en frecuencia hasta llegar a una frecuencia de comparación simple, que proporciona compatibilidad de fase para una comparación adecuada.

**Pasillos.** Debido a la considerable diferencia en frecuencia, un pasillo individual (distancia que corresponde a media longitud de onda) de DECCA es pequeño: sólo 357 m (en comparación con un pasillo de OMEGA de 14 816 m).

Existe otra característica de DECCA que permite identificar pasillos por medio de la transmisión de una señal identificadora de pasillos independiente por cada amo y esclava, una vez cada 20 segundos. Este identificador dura sólo 0.6 de segundo, pero es esencial para el uso adecuado del sistema.

**CONSOL.** Originalmente llamado "SONNE", el CONSOL fue creado por los alemanes durante la Segunda Guerra Mundial; después fue mejorado y rebautizado por los británicos. El CONSOL es una versión mejorada de RDF. Es hiperbólico en el sentido de que utiliza múltiples estaciones para generar líneas de posición hiperbólicas, pero está asociado más estrechamente con los sistemas radiales debido a que la línea base (distancia entre estaciones) en extremo corta genera un sistema de "hipérbolas colapsadas" que figuran como líneas *radiales* cuando hay alguna distancia importante a las estaciones. El principio aquí es que cuando se extiende una línea hiperbólica más allá de una distancia de aproximadamente 12 veces la longitud de la línea base, la línea se vuelve recta (llamada asíntota en términos matemáticos). Como la línea base de

CONSOL mide normalmente de 4 a 5 km, las líneas se vuelven "radiales" a una distancia relativamente corta a las estaciones.

Utilizado por aeronaves y embarcaciones de superficie, CONSOL tiene la ventaja única de que no requiere ningún equipo especializado. Las LOP de CONSOL se pueden obtener mediante el uso de una RDF o receptor estándar de comunicaciones.

**Estaciones.** CONSOL emplea estaciones equidistantes situadas en línea, donde la separación es aproximadamente tres veces la longitud de onda de la frecuencia que se transmite. Los complejos de estaciones están ubicados en las aguas europeas desde Noruega hasta España.

**Receptor.** Un receptor cualquiera de 250 a 370 kHz se puede utilizar con CONSOL. Si se utiliza un receptor de comunicaciones, el control automático de ganancia (AGC) debe estar desactivado, y resulta deseable un oscilador heterodino (BFO). El sistema produce una serie audible de puntos y rayas con información de rumbo (y por tanto LOP).

**Alcance.** Debido a la naturaleza "hiperbólica colapsada" del sistema, CONSOL tiene un alcance efectivo mínimo de 25 a 50 nmi (46 a 93 km). El alcance máximo es de 500 a 1 400 nmi (926 a 2 593 km), dependiendo de la capacidad del equipo y de las condiciones atmosféricas.

**Exactitud.** Debido a la distorsión del diagrama de radiación existe un arco inutilizable de 30° aproximadamente a cada lado de la prolongación de la línea base, con centro en la estación más cercana (fig. 52-21a). Esto significa un sector utilizable de 240° aproximadamente centrado en la bisectriz de la línea base. La exactitud es máxima en la bisectriz perpendicular de la línea base y disminuye en la proximidad de la prolongación de dicha línea. Con una exactitud máxima de 1.0 nmi, CONSOL está clasificado como sistema de mar abierto y no como costero.

**Versiónes.** En Estados Unidos se utiliza un sistema semejante en tres sitios: Nantucket Island, San Francisco y Miami. La versión estadounidense de CONSOL se denomina CONSOLAN, emplea sólo dos antenas en lugar de tres y funciona a un nivel de potencia de salida más alto, en el intervalo de 190 a 194 kHz.

La Unión Soviética tiene asimismo una versión de CONSOL llamada BPM5, que se vale de una cadena de cinco estaciones en un patrón (diagrama) cruzado. Esta modificación produce sectores de puntos y rayas más estrechos, y de este modo mejora la exactitud.

**Sistema de navegación por inercia de barcos.** El sistema de navegación por inercia de barcos (SINS) es un sistema electrónico de navegación único en varios aspectos importantes:

1. No requiere estaciones transmisoras, y por tanto no es hiperbólico ni radial.

2. Proporciona un punto de posición en vez de una serie de LOP.
3. Requiere personal especialmente cualificado de operación y mantenimiento, y es en extremo costoso.
4. Es independiente dentro de la embarcación y no requiere fuente electrónica exterior para funcionar (sin embargo, se necesitan actualizaciones periódicas de corrección, con datos que deben obtenerse de alguna fuente externa, tal como los satélites de navegación).

El SINS se desarrolló originalmente para ser usado en aeronaves y naves espaciales, y después halló aplicación en el programa de submarinos Polaris. Se ha ampliado para utilizarse en embarcaciones de superficie de la marina de Estados Unidos.

**Principios de operación.** Dos dispositivos que perciben aceleraciones respecto a la superficie de la Tierra (conocidos como acelerómetros) son estabilizados por un sistema de componentes giroscópicos para garantizar una orientación continua norte-sur y este-oeste. Los movimientos de la embarcación, y por tanto de los acelerómetros estabilizados, son captados y, mediante la segunda ley del movimiento de Newton, se integran para determinar la velocidad y distancia con respecto al tiempo. Un componente de computadora del sistema convierte las coordenadas rectangulares resultantes en sus equivalentes esféricas para compensar la curvatura terrestre. El resultado es una lectura continua de latitud y longitud, así como del apromamiento o rumbo (dirección) y la velocidad del barco.

**Errores.** La rotación diaria de la Tierra, mínima pero no obstante existente; la fricción; y la incorrecta interpretación de los efectos gravitacionales contribuyen con errores muy pequeños pero acumulativos en el sistema SINS, que se debe depurar en forma periódica. Esto se logra actualizando la posición real de la embarcación a través de alguna otra fuente de información confiable.

### LIBROS Y REVISTAS

- *Admiralty Manual of Navigation*, Vol. 1, Her Majesty's Stationery Office, Londres.
- S. F. Appleyard, *Marine Electronic Navigation*, Routledge & Kegan Paul, Londres.
- G. E. Beck, *Navigation Systems: A Survey of Modern Electronic Aids*, Van Nostrand Reinhold, Londres.
- N. Bowditch, *American Practical Navigator*, Defense Mapping Agency Hydrography Center.
- C. H. Cotter, *The Elements of Navigation and Nautical Astronomy*, Brown, Son & Ferguson, Glasgow.
- H. N. Davies, *The Navigators' Guide to Hyperbolic Navigation*, Brown, Son & Ferguson, Glasgow.
- Defense Mapping Agency, *Radio Navigational Aids*, Publication 117A & B, Defense Mapping Agency Office of Distribution Services.

- R. R. Hobbs, *Marine Navigation 2: Celestial and Electronic*, U.S. Naval Institute, Annapolis, MD.
- Institute of Navigation, *Navigation: Journal of the Institute of Navigation*, Washington, DC (trimestral).
- J. D. Lenk, *Electronic Navigation Made Easy*, John F. Rider, Nueva York.
- Litton Systems Inc., *An Introduction to OMEGA*, Litton Aero Products, Canoga Park, CA.
- E. S. Maloney, *Dutton's Navigation & Piloting*, U.S. Naval Institute, Annapolis, MD.
- A. B. Moody, *Navigation Afloat: A Manual for the Seaman*, Van Nostrand Reinhold, Nueva York.
- D. A. Paushing, *Termination of LORAN-A*, Oregon State University, Corvallis, OR.
- Time-Life Library of Boating, *Navigation*, Time-Life Books, Alexandria, VA.
- U.S. Coast Guard, *Light List CG 158*, Department of Transportation, Washington, DC.
- U.S. Coast Guard, *LORAN-C User Handbook*, U.S. Government Printing Office, Washington, DC.
- United States Naval Institute, *Proceedings*, Annapolis, MD (mensual).
- U.S. Naval Oceanographic Office, *Navigation Dictionary*, U.S. Government Printing Office, Washington, DC.



## CAPÍTULO 53

# Sistemas de aterrizaje de aeronaves

**Stephen C. Martin**

Lockheed Missiles and Space Company  
Palo Alto, California

### **53.1 Radioayuda para el aterrizaje de aeronaves**

- 53.1.1 Aterrizaje de precisión y no de precisión por instrumentos

### **53.2 Sistema de aterrizaje por instrumentos (ILS)**

- 53.2.1 Localizador
- 53.2.2 Indicador de inclinación de la trayectoria de planeo
- 53.2.3 Radiobalizas marcadoras
- 53.2.4 DME del ILS
- 53.2.5 ILS de categorías II y III
- 53.2.6 Equipo ILS de a bordo
- 53.2.7 Planes de futuro
- 53.2.8 Altímetro radioeléctrico
- 53.2.9 Sistema estándar provisional de aterrizaje por microondas (ISMLS)

### **53.3 Radar de precisión para el aterrizaje (PAR)**

### **53.4 Sistema de aterrizaje por microondas (MLS)**

- 53.4.1 Formato de las señales
- 53.4.2 Procesamiento a bordo
- 53.4.3 Exactitud de guiage
- 53.4.4 Equipo radiotelemétrico de precisión (PDME)
- 53.4.5 Equipo terrestre MLS
- 53.4.6 Equipo de a bordo

### **53.5 Sistemas de aterrizaje militares para fines especiales**

- 53.5.1 Sistemas de aterrizaje AIL
- 53.5.2 AN/SPN-42
- 53.5.3 Equipo de guiage digital de aeronaves por microondas (MADGE)
- 53.5.4 Sistema de aproximación y aterrizaje de la marina en sitios remotos (MRAALS)

### 53.1 RADIOAYUDA PARA EL ATERRIZAJE DE AERONAVES

Desde los primeros días de la aviación se reconoció con rapidez que la fase más crítica del vuelo era la de aproximación final y aterrizaje. Una vez que se facilitó la navegación en ruta entre nubes por medio de instrumentos de vuelo giroscópicos y los primeros radiofaros direccionales, se dedicó un esfuerzo considerable al desarrollo de auxiliares para guiar el aterrizaje. En la referencia 1 se hace una descripción de los primeros trabajos realizados por el gobierno de Estados Unidos en esta área. En 1929, James H. Doolittle (más tarde general) efectuó un aterrizaje totalmente a ciegas mediante el uso del radiofaro direccional de baja frecuencia como guía lateral y un altímetro sensible como referencia vertical. Estos experimentos demostraron que, para una operación confiable en la proximidad de tierra, era necesaria una señal de guiage vertical precisa. Otros trabajos realizados en Estados Unidos y Europa condujeron al desarrollo del sistema de aterrizaje por instrumentos (ILS).

Las radioayudas para el aterrizaje no suelen emplearse para guiar una aeronave hasta tocar tierra, sino más bien para orientarla en la aproximación hasta que se establece contacto visual con la pista; desde este punto se puede realizar un aterrizaje normal. La capacidad de aterrizaje confiable totalmente automático de una compañía de transporte comercial es relativamente reciente; el primer aterrizaje automático de una aeronave comercial con pasajeros fue realizado por la British Airways en 1965\*.

#### 53.1.1 Aterrizaje de precisión y no de precisión por instrumentos

En un aterrizaje no de precisión por instrumentos, el guiage lateral es provisto por uno o más de los auxiliares de radionavegación de corta distancia, que se describen en la sección 52.2 [p. ej. el radiofaro omnidireccional de muy alta frecuencia (VHF), equipo radio-telemétrico (DME), sistema de navegación aérea táctica (TACAN), radiofaros no direccionales (NDB)]. El piloto realiza descensos cronometrados a altitudes específicas después de pasar una serie de una o más posiciones fijas (marcaciones) de radio. Cada procedimiento de aterrizaje aprobado es peculiar a un aeropuerto o pista específicos y se representa en las cartas de manera individual. Las altitudes de descenso mínimas legales antes de establecer contacto visual para el aterrizaje varían en forma considerable con las condiciones del terreno, pero rara vez son menores de 500 pies sobre tierra. Los aterrizajes no de precisión

\* N. del T.: Si bien como se acaba de ver existe una clara diferencia entre la aproximación y el aterrizaje, dada la estrecha interrelación de ambas operaciones y a menos que haya la posibilidad de confusión, en lo sucesivo se utilizará el término aterrizaje en un sentido amplio que comprende tanto las maniobras de aproximación como el aterrizaje propiamente dicho, entendiendo como el momento de tomar tierra la aeronave y las maniobras asociadas inmediatamente anteriores y posteriores.

por instrumentos son realizados en forma rutinaria por aeronaves privadas, militares, corporativas y comerciales de menor tamaño en miles de aeropuertos de todo el mundo.

Un aterrizaje de precisión por instrumentos es aquel en el cual el piloto recibe orientación vertical continua por medios electrónicos. En el caso de las aeronaves comerciales existen en la actualidad tres sistemas de aterrizaje de precisión de este tipo: el sistema de aterrizaje por instrumentos (ILS) ordinario; el sistema de aterrizaje por microondas (MLS), que ahora se encuentra en desarrollo avanzado; y el radar de precisión para el aterrizaje (PAR), sistema de guiage radiotelefónico que, cuando es utilizado para fines militares, es un elemento para el aterrizaje controlado desde tierra (GCA). Las características de estos sistemas son definidas por los convenios de la International Civil Aviation Organization (ICAO).

Existe asimismo un número considerable de sistemas de aterrizaje especiales que se han desarrollado para satisfacer necesidades militares, tales como el aterrizaje de transportes de carga aéreos y la operación táctica de helicópteros y aeronaves de despegue y aterrizaje verticales (VTOL).

Para el ILS ordinario se han definido diversas categorías de servicio, como se observa en la tabla 52-1. Éstas se aplican ahora con frecuencia al funcionamiento de los sistemas de aterrizaje de precisión en general. Es claro que para la operación en condiciones de la categoría III se requiere alguna forma de capacidad de aterrizaje automático.

### 53.2 SISTEMA DE ATERRIZAJE POR INSTRUMENTOS (ILS)

El ILS ordinario consta de tres elementos: el localizador, que proporciona la orientación lateral; el indicador de la trayectoria de descenso, que da la orientación vertical; y una o más radiobalizas marcadoras (de aproximación), que definen puntos específicos a lo largo del curso. El conjunto en su totalidad define un rumbo recto, de unas 10 millas náuticas (nmi) de longi-

**Tabla 53-1. Definiciones de la ICAO<sup>a</sup> de categorías de visibilidad en el aterrizaje**

Categoría	Distancia visual de la pista (RVR)	Altura de decisión (continuar visualmente o dar la vuelta)
I	800 m	60 m (200 pies)
II	400 m	30 m (100 pies)
III(a)	200 m	0
III(b)	50 m	0
III(c)	0	0

Fuente: International Standards and Practices<sup>2</sup>.

<sup>a</sup> ICAO, International Civil Aviation Organization.

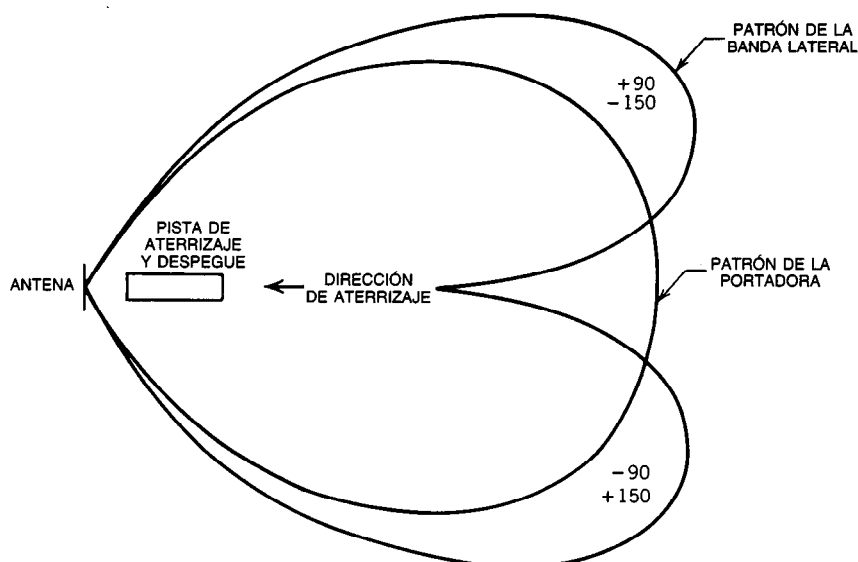


Fig. 53-1. Diagramas de radiación del localizador.

tud, que desciende a un ángulo fijo, por lo general de  $3^\circ$ , al extremo cercano de la pista. El localizador opera en uno de 40 canales de VHF entre 108.1 y 111.95 MHz [intercalados con canales omnidireccionales (VOR) de VHF]. El espaciamiento mínimo entre canales es de 50 kHz. La pendiente de planeo ocupa uno de 40 canales de ultraalta frecuencia (UHF) entre 329.15 y 334.7 MHz (con espaciamiento mínimo de 150 kHz), cada uno de los cuales está "pareado" con un canal localizador lateral (ref. 2, sec. 3). Las radiobalizas marcadoras operan a 75 MHz, igual que las radiobalizas de ruta que se describen en la sección 52.2.

Demostrada por vez primera en 1939, la configuración actual del ILS data de 1942 y se utilizó ampliamente en la Segunda Guerra Mundial. Para 1985 se esperaba que hubiera más de 1 200 instalaciones en tierra en todo el mundo, con más de 150 000 aeronaves equipadas.

### 53.2.1 Localizador

La señal del localizador es modulada en amplitud por dos señales a 90 y 150 Hz, y radia continuamente con un sistema de antenas direccionales de elementos múltiples en fase mediante el uso de polarización horizontal. La aeronave utiliza una antena receptora omnidireccional. Cuando se está "en rumbo", se reciben niveles iguales de las señales desmoduladas de 90 y 150 Hz. Cuando la aeronave está a la izquierda del rumbo, la señal de 90 Hz es más intensa; la de 150 Hz predomina cuando aquella está a la derecha del rumbo. En la figura 53-1 se presentan los diagramas de radiación típicos de los dos conjuntos de señales del localizador, las señales "portadora" y de "banda lateral". La señal portadora en realidad se somete a modulación en amplitud de doble banda lateral con ni-

veles iguales (normalmente, 20% de modulación) de las señales de 90 y 150 Hz. La señal de banda lateral contiene modulación DSB de portadora suprimida en 90 y 150 Hz. En el lado izquierdo del rumbo (parte inferior de la fig. 53-1), la modulación de 90 Hz está en fase con la de la señal portadora y la de 150 Hz está fuera de fase. A la derecha del rumbo estas fases transmitidas se invierten. Así, en el receptor de la aeronave, la onda recibida compuesta favorece la frecuencia de 90 o la de 150 Hz cuando está a la izquierda o a la derecha del rumbo, respectivamente.

La modulación transmitida y los diagramas se ajustan de modo que la diferencia en profundidad de modulación (DDM) recibida varíe linealmente con el ángulo fuera de rumbo en un intervalo aproximado de  $\pm 2.5^\circ$  (según la instalación). En los extremos de la cobertura, la DDM es  $\pm 0.175$ . El equipo de las aeronaves proporciona señales "de escala real" a la izquierda o la derecha de esta DDM. Los volúmenes de cobertura del localizador recomendados son definidos por la ICAO de la manera siguiente:

$\pm 10^\circ$ desde la línea de rumbo	distancia de 25 nmi
$10^\circ$ a $35^\circ$ desde la línea de rumbo	distancia de 17 nmi
Mayor de $35^\circ$ desde la línea de rumbo	distancia de 10 nmi
Cobertura vertical	$0^\circ$ a $7^\circ$ sobre el horizonte

La intensidad de campo mínima dentro de este volumen debe ser de  $100 \mu\text{V/m}$ . Fuera de la región de rumbo lineal ( $\pm 2.5^\circ$ ), pero dentro de la región de cobertura, la aeronave debe recibir esencialmente indicación de escala real con percepción correcta, que se conoce como señal "de franqueo". La potencia radiada



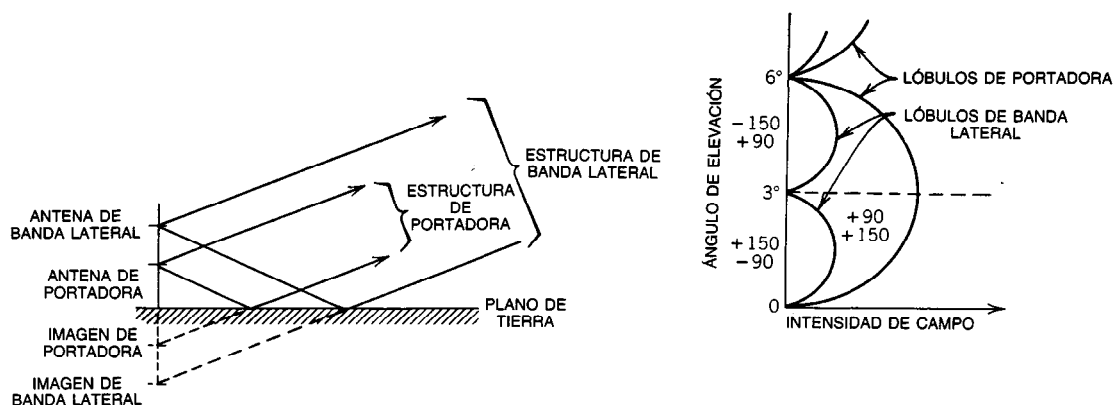


Fig. 53-2. Formación de la trayectoria de descenso (planeo) del sistema de "referencia cero".

por el localizador está en el intervalo de 10 a 15 W. Para la identificación, se transmiten en forma simultánea tonos de 1 020 Hz en clave Morse con una profundidad de modulación del 10%. En algunas partes, también se pueden superponer transmisiones de habla de la torre de control.

Se utilizan muchos tipos diferentes de arreglo (red) de antenas de localizador y de elementos del arreglo. El localizador estándar en Estados Unidos tiene ocho elementos en un arreglo lineal, que se extiende 14 m. Los detalles del espaciamiento y puesta en fase de las señales se exponen en la referencia 3. El arreglo se monta unos 300 m más allá del extremo de parada de la pista que atiende, perpendicular a la línea central extendida. Los elementos de la antena pueden ser cuadros Alford omnidireccionales (ref. 6, sec. 52.2) o tipos direccionales cuando se desea minimizar la radiación de lóbulo lateral. Entre estos últimos se encuentran el elemento de cuadro más dipolo plegado (anillo V)<sup>5</sup> y la antena de ondas progresivas<sup>6</sup>. El elemento dipolo plegado periódico logarítmico<sup>4</sup> es otro método que tiene las ventajas de altura reducida (2.1 m) y resistencia a los efectos de la formación de capas de hielo.

El uso de elementos que tienen individualmente un patrón omnidireccional presenta la ventaja de que puede existir un rumbo posterior utilizable, recíproco a la dirección de aterrizaje principal. Esto puede proporcionar guiaje de sobrevuelo y una señal de orientación para aterrizajes en dirección opuesta, en cuyo caso se invierte el sentido de la información de lado de rumbo izquierdo y derecho.

Por otra parte, esta radiación de ángulo amplio puede propiciar graves desviaciones u "ondeo" de rumbo provocado por reflexiones de trayectoria múltiple en edificios o terreno más elevado paralelo a la trayectoria de aterrizaje.

Los requisitos de la ICAO<sup>2</sup> son que la excursión (oscilación)  $2\sigma$  respecto al curso medio del localizador no debe exceder de  $0.36^\circ$  más allá de 4 nmi, y decrecer a  $0.06^\circ$ ,  $2\sigma$ , cuando esté a 1 050 m del umbral de la pista. En sitios sujetos al efecto de trayectoria múltiple, se utilizan antenas localizadoras de mucha mayor abertu-

ra para producir un diagrama de radiación más estrecho. En un arreglo típico se emplean 14 elementos en un ancho de 26 m<sup>5</sup>. Entre otros tipos de localizador de abertura amplia se cuentan los arreglos reflectores parabólicos y guías de ondas ranuradas hasta de 60 m de longitud<sup>3</sup>. La estrechez del diagrama de radiación a menudo se traduce en que las señales de franqueo adecuadas no se suministran en ángulos abiertos. Para este fin, a menudo se incluye un arreglo de abertura estrecha independiente que radia señales de guiaje en una frecuencia aproximada de 8 kHz desplazadas de las señales principales. Cuando la aeronave está próxima a la trayectoria de ruta, las señales de alta exactitud tienen mayor energía y capturan el receptor de la nave. A ángulos abiertos es la señal de franqueo la que predomina y la que suministra información de percepción correcta.

### 53.2.2 Indicador de inclinación de la trayectoria de planeo

El indicador de inclinación de la trayectoria de planeo de UHF utiliza modulaciones de amplitud de 90 y 150 Hz, como el localizador. El sistema de antena tiene una directividad vertical tal que la modulación de 90 Hz es más intensa cuando es "demasiado alta", y la de 150 Hz es más intensa cuando es "demasiado baja". Se utiliza polarización horizontal. La región lineal se extiende  $0.7^\circ$  por encima y por debajo de la trayectoria de planeo nominal, que suele fijarse en  $3^\circ$  arriba de la horizontal. En los extremos de la región lineal, la diferencia en profundidad de modulación (DDM) es  $\pm 0.175$ . La cobertura horizontal es un mínimo de  $\pm 8^\circ$  respecto a la línea central, para una distancia de 10 nmi. Dentro de esta área de cobertura la intensidad de campo mínima debe ser de  $400 \mu\text{V/m}$ . La potencia del transmisor de la pendiente de la trayectoria de descenso es de unos 4 W.

Casi todos los sistemas de antena de la pendiente de la trayectoria de planeo son del tipo imagen. En la figura 53-2 se ilustra la formación del patrón en el caso del sistema más común de "referencia cero". Se montan dos antenas dipolares horizontales sobre un poste

vertical. La inferior transmite la señal portadora, y la superior, que está dos veces más elevada, transmite únicamente las señales de banda lateral. Cada antena, junto con su imagen en el plano de tierra formado por el piso, genera una estructura de lóbulo vertical, como se muestra en el lado derecho de la figura 53-2. La estructura de lóbulo de la antena de banda lateral es dos veces más delgada que la de la antena de portadora. Para la frecuencia de la pendiente de la trayectoria de planeo de 330 MHz, las alturas aproximadas de las antenas de 4.25 y 8.5 m producen una cresta en el diagrama de la portadora y un punto nulo en el diagrama de la banda lateral al ángulo de desviación deseado de  $3^\circ$ . La reflexión en tierra hace que se invierta el ajuste de fase de las señales de banda lateral del segundo lóbulo respecto a las del primero, produciendo la percepción correcta en torno al punto nulo en ruta. Como la estructura de lóbulo prosigue en ángulos de elevación mayores, en  $15^\circ$  figura otra pendiente de trayectoria de planeo falsa estable. En la práctica esto se evita interceptando desde abajo la trayectoria de planeo correcta en vuelo a nivel. El poste de la antena se coloca unos 120 m al lado de la pista y 300 m más allá del extremo de aterrizaje, de manera que el centro de la trayectoria de planeo tiene una elevación de  $15 \pm 3$  m a través del extremo de la pista.

Las especificaciones de la ICAO requieren que las desviaciones del rumbo no excedan de  $0.15^\circ$  ( $2\sigma$ ) a 4.5 nmi y decrezcan a  $0.1^\circ$  ( $2\sigma$ ) dentro de los 1 050 m previos al toque de tierra. El arreglo de la imagen requiere que todo el terreno de la primera zona de Fresnel (sec. 48.2) sea uniforme y plano, para producir diagramas de radiación correctos. Esto es muy difícil de lograr en algunos campos de aviación. El sistema de referencia de banda lateral<sup>7</sup> es un arreglo de tipo de imagen diferente, menos sensible a los efectos del terreno. En sitios con pendiente positiva frente a la antena de la pendiente de la trayectoria de planeo, el arreglo de "efecto de captura" puede mejorar los resultados<sup>5</sup>. En esta configuración, los diagramas de radiación de las antenas independientes se conforman con el uso de pantallas reflectoras, de manera que los diagramas resultantes se inclinan sobre el horizonte. Con un tercer elemento de antena, una señal de franqueo de baja potencia, que consta sólo de información de "subida" (ascenso de la nariz), se transmite con un diagrama amplio en una frecuencia portadora de unos 8 kHz, desplazada de la radiación primaria. Cerca del "rumbo" la señal principal es más intensa, pero debajo de la región de rumbo lineal la señal de franqueo es más intensa y captura el receptor de la aeronave, produciendo una instrucción de subida intensa.

En sitios especialmente difíciles, se han utilizado antenas no de imagen de pendiente de trayectoria de planeo. Se han probado guías de ondas verticales ranuradas de 330 MHz hasta de 35 m de longitud<sup>5</sup>, inclinadas  $3^\circ$  de la vertical para generar la pendiente requerida del haz. Una antena de 20 m de este tipo está instalada en LaGuardia, Nueva York, donde el plano de tierra es en realidad la superficie del mar y no puede emplearse para reflejar una imagen debido a variaciones de 3 m de la marea.

Se ha investigado intensamente sobre configuraciones no de imagen de radiación longitudinal<sup>6,8</sup>. Una configuración en uso tiene dos fuentes de radiación separadas más o menos 200 m, cada una de ellas desviada 60 m de la línea central de la pista y cerca de 1.25 m sobre tierra. Cada una de ellas está alimentada con una portadora y ambas bandas laterales en cuadratura de fase.

A pesar de estos sistemas de pendiente de trayectoria de planeo diferentes, en muchos aeropuertos pequeños y en los de regiones montañosas aún no es factible aplicar un ILS de pendiente de trayectoria de planeo aceptable.

### 53.2.3 Radiobalizas marcadoras

Se utilizan radiobalizas marcadoras de 75 MHz de baja potencia (sec. 52.2) con el ILS para definir puntos específicos a lo largo del rumbo. Cuando la aeronave pasa sobre la radiobaliza de radiación vertical, el piloto recibe unos cuantos segundos de indicación acústica y visual. Una radiobaliza marcadora exterior está instalada a una distancia de 3.5 a 6 nmi del extremo de la pista, por lo general en el punto donde la porción de vuelo horizontal del procedimiento de aterrizaje designado intercepta la pendiente de la trayectoria de planeo. El transmisor marcador exterior está modulado en amplitud al 95% con una serie de rayas de tono de 400 Hz, a razón de dos rayas por segundo. Una radiobaliza marcadora intermedia está situada a unos 1 050 m del extremo de la pista, en el punto donde la pendiente de la trayectoria de planeo está 60 m sobre tierra (altura de decisión de los aterrizajes de categoría I). Su transmisor se modula con puntos y rayas alternativos mediante el uso de un tono de 1 300 Hz. En pistas certificadas para aterrizajes de categorías II y III, también se instala una radiobaliza marcadora interna. Ésta se encuentra entre 75 y 450 m del extremo de la pista, en el punto donde la altura de la pendiente de la trayectoria de planeo es de 30 m (altura de decisión de categoría II). Se modula con puntos a razón de seis por segundo mediante el uso de un tono de 3 000 Hz.

Los transmisores de radiobalizas marcadoras ILS emplean antenas verticales de haz en abanico con polarización horizontal, y radian menos de 2.5 W. En la figura 53-3 se muestra la cobertura del localizador, el transmisor de la pendiente de la trayectoria de planeo y dos diagramas de señales de la radiobaliza marcadora.

### 53.2.4 DME del ILS

En sitios donde la instalación de radiobalizas marcadoras no resulta práctica (como en el caso de aproximaciones sobre el agua), suele instalarse equipo radiotelemétrico (DME) (que se describe en la sección 52.2). El transpondedor DME está colocado con la antena localizadora y opera en un canal que está "pareado" con el de la frecuencia VHF del localizador<sup>2</sup>. Algunos equipos DME de aviones están diseñados para reconocer éste como un canal asignado de ILS, y por

tanto conservarán una velocidad de interrogación relativamente elevada ( $\approx 50$  a  $100/s$ ) en el modo de seguimiento. Esto proporciona datos más uniformes ("suaves") al piloto automático.

### 53.2.5 ILS de categorías II y III

Casi todas las instalaciones de ILS proporcionan servicio a la categoría I estándar (altura de decisión de 60 m). En la mayor parte de los aeropuertos de Estados Unidos y Europa, un ILS de categoría II (altura de decisión de 30 m) sirve cuando menos a una pista de aterrizaje. El servicio de categoría III (altura de decisión 0) sigue siendo relativamente raro. En Estados Unidos hay poco más de media docena de sistemas de esta categoría en servicio actual. En Europa el avance es más rápido, y se ha obtenido experiencia considerable con operaciones de categoría III completamente automáticas<sup>9</sup>. La más reciente generación de aeronaves para el transporte aéreo se está produciendo con los sistemas de control de vuelo especiales necesarios para la operación de categoría III incorporados como equipo estándar (p. ej., L-1011, Concorde, A-300, Boeing 757).

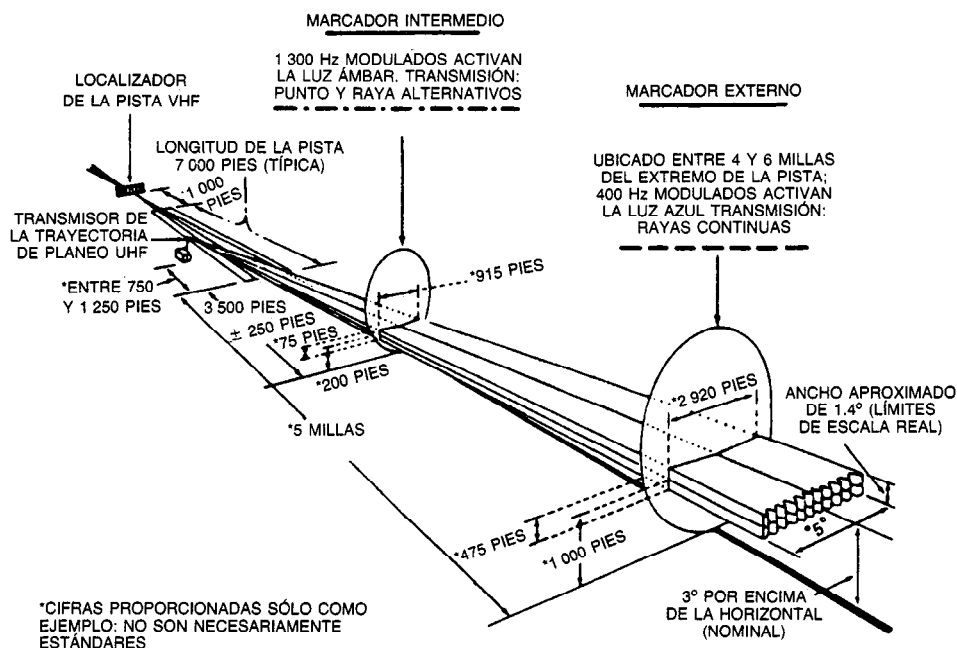
En muchos casos la diferencia entre una instalación ILS de tierra certificada para operación de categoría II o III no radica en la calidad inherente de guiaje, sino más bien en la existencia y la intensidad de uso de monitores y de equipo de transmisión duplicado. Éstos son necesarios para asegurar que se pueda hacer una conmutación casi instantánea en caso de que se detecte una señal del guiaje fuera de tolerancia.

### 53.2.6 Equipo ILS de a bordo

El localizador de aeronaves y los receptores de la pendiente de la trayectoria de planeo sintonizan cada uno el par de canales seleccionado, desmodulan en amplitud las señales y comparan por separado las amplitudes de modulación de 90 y 150 Hz. Debe tenerse cuidado para mantener un equilibrio de ganancia entre los dos canales de tono y garantizar que la diferencia cero en profundidad de modulación dé una señal de "en rumbo".

En aeronaves de menor tamaño y en los primeros tipos de equipo de líneas aéreas, las funciones del localizador se incorporan dentro del receptor VOR (sec. 52.2). La antena VOR sirve en forma análoga para recibir las señales del localizador. En este caso se utiliza un receptor de la pendiente de la trayectoria de planeo independiente, con la sintonía esclavizada a la del receptor VOR. La antena de la pendiente de la trayectoria de planeo es un dipolo plegado de UHF en forma de U, montado en la nariz del avión, o una antena de cavidad al ras. En el equipo moderno de las líneas aéreas se emplea un receptor localizador separado del VOR, combinado con el sistema de pendiente de trayectoria de planeo en una unidad "receptora ILS" común. Asimismo se incluye una antena localizadora separada, que no se comparte con el VOR.

Las normas de la ICAO (ref. 2, parte 1, anexo C) exigen selectividades de 60 dB a  $\pm 50$  kHz para receptores localizadores y de 60 dB a  $\pm 300$  kHz para receptores de pendiente de trayectoria de planeo. El Radio Technical Committee for Aeronautics (RTCA) ha establecido otras normas de funcionamiento mínimo<sup>10</sup>.



**Fig. 53-3.** Posiciones del localizador, sistema de pendiente de la trayectoria de descenso (planeo) y antena de la radiobaliza marcadora. Los radiofaros de brújula están colocados a veces con marcadores exteriores e intermedios.



Fig. 53-4. Indicaciones del indicador de intersección del sistema de aterrizaje por instrumentos (ILS).

Las sensibilidades típicas son  $3 \mu\text{V}$  para el receptor localizador y  $5 \mu\text{V}$  para el receptor de pendiente de trayectoria de planeo. Un receptor ILS completo de una línea aérea ocupa comúnmente  $5\,500 \text{ cm}^3$ , pesa  $5.5 \text{ kg}$  y consume  $26 \text{ W}$ . El receptor de pendiente de la trayectoria de planeo independiente que se emplea con el receptor (localizador) VOR/LOC en una aeronave de menor tamaño ocupa unos  $1\,400 \text{ cm}^3$ , pesa  $1 \text{ kg}$  y consume  $7 \text{ W}$ .

La visualización en la cabina del ILS más simple es el indicador de intersección de dos agujas (fig. 53-4). Este tipo de visualización suele combinarse con el indicador de desviación de rumbo (CDI) (sec. 52.2), pero esta forma de indicador de intersección básico suele encontrarse sólo en aeronaves civiles de menor tamaño y en algunas instalaciones militares. En aeronaves de líneas aéreas y comerciales la señal localizadora se conmuta a la barra de desviación del rumbo del indicador de la situación horizontal (HSI), como se muestra en la sección 52.2. Después, una escala del apuntador vertical adyacente muestra la desviación respecto de la pendiente de la trayectoria de planeo. A menudo también se incorporan apuntadores de error laterales y verticales independientes en el visualizador del indicador de la altura del piloto (horizonte artificial). Para operaciones comerciales y de líneas aéreas, en casi todos los aterrizajes con ILS los aviones se hacen descender hasta la altura de decisión en el modo acoplado al piloto automático, y los indicadores de desviación sólo se utilizan para la monitorización del piloto.

### 53.2.7 Planes de futuro

El sistema ILS ordinario, por convenio internacional, será mantenido como sistema estándar por lo menos hasta 1995. Será sustituido finalmente por el sistema de aterrizaje por microondas (MLS) de haz explorador, que se describe más adelante.

### 53.2.8 Altimetro radioeléctrico

El altímetro radioeléctrico o radioaltímetro es un radar dirigido hacia abajo de baja potencia utilizado por aeronaves civiles para determinar la altura sobre tierra en las fases finales de aproximación para el aterrizaje. El equipo para aeronaves debe estar certificado para hacer aterrizajes en condiciones de categoría II o III. Los altímetros radioeléctricos civiles operan en la banda de  $4.2$  a  $4.4 \text{ GHz}$ . El RTCA establece normas de

funcionamiento<sup>12</sup>. Se utilizan sistemas de onda continua (CW) modulada y de pulsos. En los tipos de CW modulada se utiliza modulación triangular o sinusoidal, midiendo la altura por medio de la diferencia de frecuencia entre señales transmitidas y recibidas (ref. 13, sec. 3.3). Una desviación de modulación típica es de  $100 \text{ MHz}$  de cresta a cresta a una razón de  $100 \text{ Hz}$ .

Los altímetros de radar civiles operan hasta una altura máxima aproximada de  $750 \text{ m}$  ( $2\,500$  pies), pero es de mayor interés la información por debajo de los  $150 \text{ m}$  ( $500$  pies). El visor de radioaltitud del piloto tiene una escala extendida en esta región que muestra la altura a una resolución de más o menos  $3 \text{ m}$  ( $10$  pies). El altímetro del radar puede prefijarse para activar una señal de advertencia visible y audible cuando la aeronave desciende a través de la altura de decisión en la categoría de aterrizaje en que se vuela. Para los aterrizajes automáticos de la categoría III, el guiage vertical de la aeronave se transfiere progresivamente de sistema de pendiente de trayectoria de planeo ILS a la señal del radioaltímetro cuando la nave desciende a los últimos  $30 \text{ m}$  ( $100$  pies) de la superficie.

Casi todos los altímetros del tipo de pulsos utilizan una sola antena para recibir y transmitir, transmiten de  $5$  a  $10 \text{ W}$  de potencia pulsátil pico y tienen un alcance mínimo (fijado por consideraciones de amplitud del pulso) de alrededor de  $12 \text{ m}$ . Los altímetros de CW-FM utilizan dos antenas individuales, radian unos  $400 \text{ mW}$  y se pueden emplear hasta una altura cero. La exactitud típica es  $\pm 0.5 \text{ m}$  o el  $2\%$  de la altura a menos de  $150 \text{ m}$ . Un sistema de pulsos de una aeronave pequeña típicamente ocupa  $1\,400 \text{ cm}^3$ , pesa  $1 \text{ kg}$  y consume  $6 \text{ W}$ . Para una unidad del tipo CW-FM de línea aérea, estas cifras son  $9\,000 \text{ cm}^3$ ,  $5.5 \text{ kg}$  y  $50 \text{ W}$ .

Los altímetros de radar militares pueden operar en frecuencias diferentes de la banda civil de  $4\,300 \text{ MHz}$ . Al igual que para el aterrizaje, se utilizan para realizar la medición en ruta en la altura sobre el terreno, hasta altitudes de  $6\,000 \text{ m}$  o más. En el caso de una aeronave equipada con un sistema de navegación Doppler (sec. 52.2), la radioaltura a menudo puede determinarse directamente a partir de los datos que se utilizan para la navegación.

### 53.2.9 Sistema estándar provisional de aterrizaje por microondas (ISMLS)

Se trata de un sistema de microondas ( $5 \text{ GHz}$ ) en el que se utilizan haces de exploración en escalones para

el guiage de acimut y vertical, y que hasta el momento en que esto se escribe se ha instalado en unos 12 sitios de Estados Unidos desde 1975. Fue creado para proporcionar guiage de tipo ILS en lugares donde la instalación del ILS ordinario era demasiado complicada. El acimut de  $\pm 40^\circ$  está cubierto por ocho haces conmutados, y la elevación de 0 a  $6^\circ$  con cuatro haces. El régimen de trabajo aproximado de cada conjunto es a razón de 600 Hz. Cada haz se modula en ancho de pulso con valores de muestra de una onda de forma compuesta de 90/150 Hz, con razones de amplitud relativas apropiadas para la distancia entre el haz y la línea central o trayectoria de planeo de  $3^\circ$ . La amplitud del receptor de la aeronave detecta las frecuencias de señal de acimut y vertical adecuadas y obtiene dos conjuntos de señales de 90/150 Hz proporcionales que son función del ángulo de error, como en el caso del ILS. De hecho, en el receptor de la aeronave se utilizan localizadores-receptores de la pendiente de la trayectoria de planeo de UHF-VHF ordinarios [con mínimas modificaciones del control de ganancia automático (AGC)], precedidos de una etapa única de conversión descendente a partir de los 5 GHz. Las dos señales de 5 GHz se transmiten con el desnivel adecuado ( $\approx 200$  MHz), de manera que las señales convertidas descendentes de "localizador" y de "receptor de la pendiente de la trayectoria de planeo" tengan uno de los pareamientos de frecuencia ILS estándar.

Este formato de señales es totalmente diferente del propio del MLS de la ICAO, que se describe en la sección 53.4. Ya se ha suspendido un desarrollo más a fondo de ISMLS en favor de este MLS "estándar".

### 53.3 RADAR DE PRECISIÓN PARA EL ATERRIZAJE (PAR)

El PAR es un auxiliar de guiage basado en tierra por medio del cual una aeronave puede efectuar aproximaciones hasta baja altura sin llevar equipo especial alguno que no sea un receptor de radio. Un radar de tierra de precisión sigue cada aeronave que llega, y en pantallas CRT de elevación y acimut simultáneas indica la posición de la nave en términos de ángulos de pendiente de la trayectoria de planeo y de acimut deseados. Un operador capacitado, al observar estas pantallas, radia instrucciones habladas de corrección al piloto.

La ICAO ha fijado especificaciones de funcionamiento del PAR para uso de la aviación civil (ref. 2, sec. 3.2). Éstas exigen que el radar sea capaz de seguir un objetivo de  $15 \text{ m}^2$  (equivalente a una pequeña aeronave bimotor) a una distancia de 9 nmi. La exactitud de medición de la posición debe ser de 9 m en sentido lateral, 6 m verticalmente y 30 m en alcance. El PAR proporciona guiage hasta mínimos de categoría I de ILS (altura de decisión de 60 m). Sin embargo, ahora apenas se utiliza en las aeronaves civiles. El PAR se utilizó alguna vez como apoyo en muchos aeropuertos grandes para monitorizar aeronaves que hacían aterrizajes ILS, pero la mayor parte de estas instalaciones ya se han retirado.

Los militares han seguido utilizando el PAR desde su desarrollo en la Segunda Guerra Mundial<sup>114</sup>. Una instalación militar que consta de un radar de vigilancia para adquisición y del PAR para el guiage final se conoce como sistema de aproximación-aterrizaje controlado desde tierra (GCA). Este sistema es mucho más adecuado para bases avanzadas que el ILS, que requiere una extensa preparación y nivelación del sitio de ubicación. En muchas circunstancias se utiliza un PAR para guiar aeronaves militares hasta mínimos equivalentes de ILS de categoría II (altura de decisión de 30 m).

El PAR opera comúnmente en 9 a 10 GHz. En muchos sistemas en uso se emplean dos antenas ortogonales de haz en abanico de exploración mecánica para realizar mediciones separadas en acimut y elevación. Un ejemplo de sistema moderno es el radar de precisión AN/TPN-19. Éste utiliza una sola antena reflectora de 824 elementos montados en fase, que mide 2.8 m de ancho por 3.55 m de alto, para generar un haz puntual de  $1.4^\circ$  en acimut por  $0.75^\circ$  en elevación. Puede operarse en exploración continua para búsqueda o como radar de persecución monopulso sobre un sector de  $\pm 10^\circ$  de acimut por  $15^\circ$  de elevación. Las velocidades de transferencia de datos son de 22 exploraciones por segundo en búsqueda y 2 por segundo en seguimiento: Con una potencia pico de 320 kW, el intervalo de operación con una intensidad de lluvia de 50 mm/h es de 15 nmi. La exactitud del seguimiento angular es  $0.143^\circ$  en acimut y  $0.072^\circ$  en elevación. Mediante el uso de un pulso de  $1 \mu\text{s}$  con 120 MHz de compresión de la frecuencia del tono agudo (compresión de pulso), se ha informado una resolución de distancia de 1.5 m.

La marina de Estados Unidos utiliza un sistema PAR en el cual las instrucciones habladas del operador son substituidas por un enlace de datos de la superficie al aire, el AN/SPN-42, para aterrizajes de transportes aéreos. Esto se analiza más adelante.

### 53.4 SISTEMA DE ATERRIZAJE POR MICROONDAS (MLS)

El MLS ha sido adoptado por la ICAO como el sistema que substituirá finalmente al ILS ordinario después de 1995. La coexistencia de ambos sistemas se pronostica por un tiempo considerable. El MLS tiene muchas ventajas operacionales sobre el ILS ordinario, como:

1. Guiage para aproximaciones de precisión por instrumentos en sitios donde el terreno es demasiado irregular para la instalación de ILS.
2. Guiage para perfiles de aproximación curvos a fin de reducir el uso del espacio aéreo y evitar áreas sensibles al ruido en tierra.
3. Ángulos de trayectoria de descenso (planeo) variables para ajustarse al funcionamiento de cada aeronave.
4. Mayor número de canales en operación.
5. Menores costos de instalación y mantenimiento de equipo terrestre.

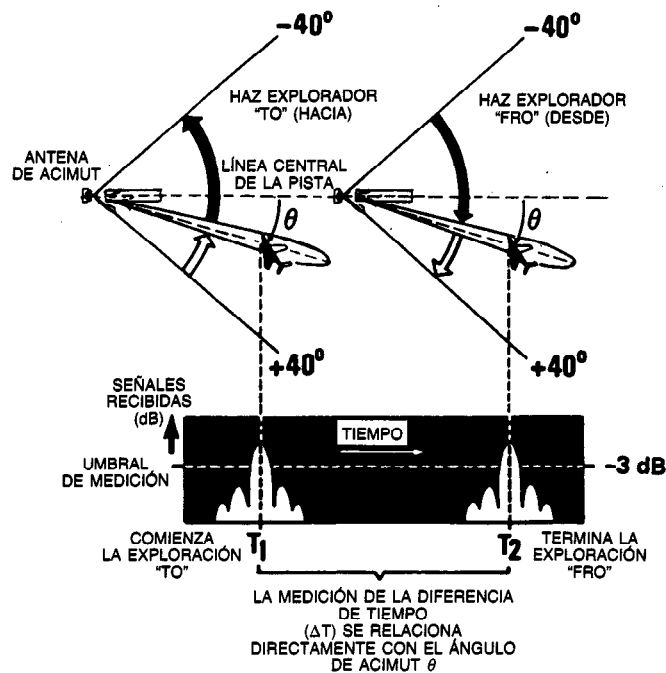


Fig. 53-5. Medición de ángulos del sistema de aterrizaje por microondas (MLS). Tomado de Cox y Shirey<sup>15</sup>, con permiso de IEEE.

En Estados Unidos, Inglaterra, Francia, Japón, Australia y la Unión Soviética se están probando prototipos de equipo MLS terrestre y de a bordo. Estados Unidos ha realizado un programa de pruebas de servicio y evaluación (STEP) mediante el uso de estaciones en tierra en Atlantic City, Filadelfia y Washington, D.C.<sup>15</sup>

Los servicios militares de Estados Unidos están apoyando el desarrollo de un sistema táctico conjunto de aterrizaje por microondas (JTMLS) compatible. Este programa producirá una estación terrestre transportable de mochila y medios electrónicos de aeronáutica (aviónica) adecuados para probar los procedimientos conjuntos de ejército, armada, marina mercante y fuerza aérea, mediante el uso de aviones de ala fija y helicópteros.

El esquema de guiado angular MLS se conoce como esquema de haz de exploración con referencia de tiempo (TRSB). Esta técnica fue seleccionada por la ICAO en 1978 después del estudio y evaluación de varios sistemas de guiado competitivos. Estos eran el esquema de haz de exploración (Doppler) de antena conmutada, la multilateralización mediante equipo radiotelemétrico (DME) y métodos de haces de exploración codificados en frecuencia.

Con el concepto TRSB, unos haces en abanico de exploración ortogonal proporcionan datos angulares. La estrecha amplitud de cada haz en su dimensión crítica hace posible que el procesador de la aeronave realice una discriminación importante entre las señales correctas y las generadas por reflexiones de trayectoria múltiple en estructuras o tierra.<sup>16</sup>

Esta capacidad permitirá utilizar el MLS en sitios de aterrizaje en los cuales no sea aplicable el ILS ordinario.

Las funciones de guiado angular y de transmisión de datos en un sitio dado se multicanalizan en el tiempo en un solo canal de frecuencia en la banda C. Existen 200 canales asignados a intervalos de 300 kHz, desde 5 031.0 hasta 5 090.7 MHz. Cuando se incluyen funciones de radioteleetría en un sitio de MLS, éstas emplean las frecuencias y formatos DME de banda L ordinarios, salvo que el espaciamiento y procesamiento de los pulsos se modifican para corta distancia a fin de mejorar la precisión. La ICAO ha publicado especificaciones provisionales de las funciones de guiado y datos angulares (ref. 2, parte I, anexo I). En 1984 se adoptaron especificaciones similares para el DME de precisión.

El ILS ordinario genera una sola trayectoria de guiado recta, fija en acimut y elevación. Por otra parte, el MLS transmite "señales en el espacio" a partir de las cuales la aeronave que se aproxima puede medir su acimut y elevación desde la línea central de la pista sobre un volumen que (comúnmente) se extenderá  $\pm 40^\circ$  en acimut, de 0 a  $10^\circ$  en elevación y 20 nmi a partir del toque de la pista. Dentro de este volumen el equipo del avión puede determinar sus desviaciones respecto de cualquier trayectoria de aproximación-aterrizaje preseleccionada (que podría ser curva) y cualquier ángulo de descenso preseleccionado. Los datos que transmite la estación MLS de tierra proporcionan (entre otras cosas) una advertencia de obstrucciones en acimut y elevación.

## 53.4.1 Formato de las señales

Los haces en abanico de acimut y elevación exploran electrónicamente a velocidades controladas con precisión. La figura 53-5 ilustra el principio de la medición de acimut. El estrecho haz en abanico orientado en sentido vertical realiza un par de exploraciones laterales de un lado a otro del área de cobertura, llamadas exploraciones "TO" y "FRO" (hacia y desde). El acimut de la aeronave,  $\theta$ , se determina a partir de la diferencia de tiempo entre la detección del paso de estos dos haces:

$$\theta = \frac{T_0 - t}{K}$$

donde  $\theta$  = ángulo requerido

$t$  = intervalo medido,  $\mu s$

$T_0$  = constante conocida que corresponde al retardo cuando  $\theta = 0$ ,  $\mu s$

$K$  = constante de escalación conocida,  $\mu s/^\circ$

El ángulo de elevación se define en forma análoga mediante el uso de un haz en abanico lateral, que realiza exploraciones verticales "hacia" y "desde". Para el acimut,  $T_0$  es 6 800.0  $\mu s$ , y para la elevación, 3 466.7  $\mu s$ . En ambos casos  $K$  es 100  $\mu s/^\circ$ .

Las funciones de elevación y acimut constituyen el formato básico del MLS. Se pueden agregar otras capacidades sobre una base modular, como un haz de guiage lateral para "aproximación frustrada" o de contraacimut, y un haz de guiage de elevación de alta precisión independiente que cubren la región donde la aeronave comienza su "enderezamiento" justo antes de aterrizar. Cada función se transmite en serie en el canal de frecuencia asignado, como se ilustra en la figura 53-6. Cada función individual consta de un preámbulo seguido de un par de búsqueda TO-FRO (hacia-desde). El preámbulo está formado por 400  $\mu s$  de portadora para la adquisición, seguidos de una palabra de sincronización de 5 bits, 5 bits que identifican la función y 2 bits de paridad extra, para el acimut, un bit adicional para activar y desactivar en la aeronave una señal de identidad en clave Morse audible. Estos datos se transmiten mediante manipulación por defasamien-

to diferencial a 15 kb/s. Cada función es procesada en forma independiente por el receptor de la aeronave. La transmisión de cada función de guiage y datos puede intercalarse por tanto con otras funciones. Las velocidades de repetición mínimas de cada una de las funciones, como lo especifica la ICAO<sup>2</sup>, son:

Función	Velocidad de repetición (Exploraciones TO/FRO por segundo)
Acimut	13.5
Elevación	40.5
Contraacimut	6.75
Elevación de enderezamiento	40.5

En el formato se toman medidas para la futura adición de una función de acimut de 360° independiente. Entre los formatos de exploración se cuentan pulsos de prueba y, para la exploración del acimut, pulsos de indicación de fuera de cobertura (OCI). Estos son equivalentes a los patrones de franqueo del ILS ordinario, y sus amplitudes recibidas relativas se utilizan para advertir a una aeronave que se encuentra fuera de la región de guiage angular proporcional.

La transmisión de datos de tierra a aire es parte inherente de la señal MLS. Estos datos se transmiten en bloques en el formato de tiempo compartido mediante el uso de la manipulación por defasamiento diferencial (DPSK), y se radian desde el sitio de acimut básico empleando un diagrama de antena que llene el volumen de cobertura de guiage. Los formatos de estos datos todavía no están definidos por completo. Los datos básicos se utilizarán por todas las aeronaves. Entre ellos se cuentan:

- Categoría de funcionamiento.
- Condición del equipo.
- Anchos de haces de antena.
- Límites de cobertura de acimut.
- Ángulo de trayectoria de descenso (planeo) seleccionable mínimo.
- Número de canal DME asociado (si se usa).

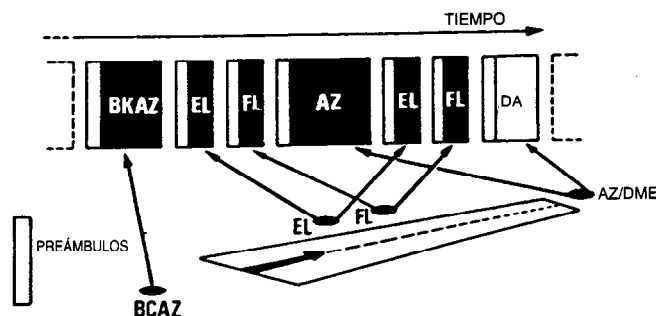


Fig. 53-6. Formato de señales del sistema de aterrizaje por microondas (MLS). BKAZ (BCAZ) = contraacimut; EL = elevación; FL = enderezamiento; AZ = acimut; DME = equipo radiotelemétrico; DA = datos auxiliares. Tomado de Cox y Shirey<sup>15</sup>, con permiso de IEEE.

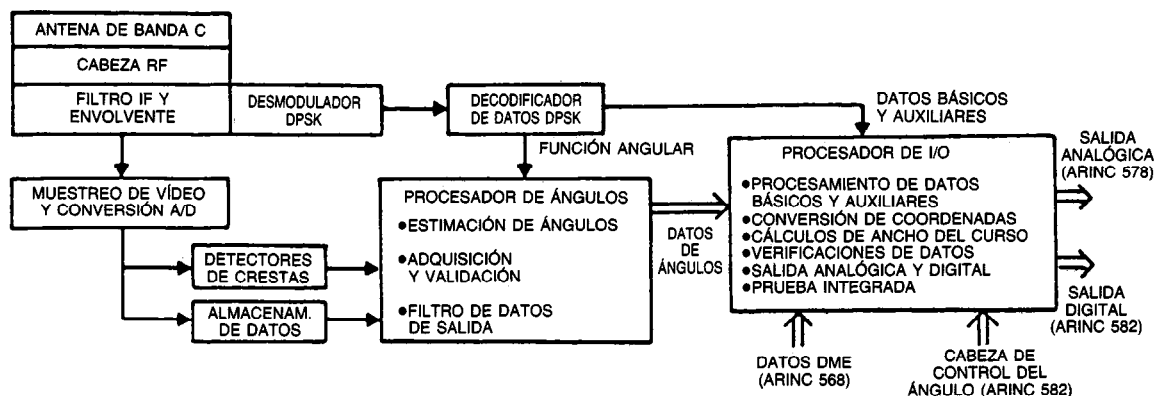


Fig. 53-7. Arquitectura de procesamiento del receptor de ángulos. DPSK = manipulación por defasamiento diferencial; DME = equipo radiotelemétrico. Tomado de Kelly y Skudrna<sup>17</sup>, con permiso de IEEE.

Después de la transmisión de datos básicos, se transmitirán datos auxiliares para su uso por aeronaves equipadas en forma adecuada, tal como para el envío de mensajes en caracteres alfanuméricos.

#### 53.4.2 Procesamiento a bordo

La figura 53-7 ilustra una arquitectura típica<sup>17</sup> para el procesamiento a bordo de ángulos y datos en MLS. Se implementa utilizando dos microprocesadores, uno para la medición de ángulos y el otro para emplear esta información, más los datos del sitio de ubicación, a fin de generar salidas de posición y gobierno. Salidas análogas se envían a los visores del tipo de indicador de intersección, y salidas digitales a la interfaz con un piloto automático. El procesador de I/O recibe asimismo entradas desde el equipo radiotelemétrico (DME) y desde la unidad de control, en la cual el piloto selecciona el seguimiento de vuelo y ángulo de descenso.

Es posible eliminar por filtrado en forma considerable el efecto de trayectoria múltiple en el procesamiento de mediciones de ángulos. El procesador de señales de compuerta dividida (SGSP)<sup>17</sup> determina el tiempo de llegada de las crestas de los haces TO y FRO, diferenciado esencialmente la respuesta de vídeo e identificando el cruce por cero de esta derivada. La posición de este cruce por cero es relativamente insensible a la distorsión por trayectoria múltiple de la forma global del pulso. El SGSP realiza interpolaciones sucesivas sobre secuencias de muestras de cada respuesta de pulso, a fin de determinar el cruce por cero en una forma que sea menos sensible al ruido que la diferenciación ordinaria.

Cuando se proporciona una señal de guiage de elevación de enderezamiento independiente para el aterrizaje automático, se aplica una técnica conocida como procesamiento de un solo flanco (SEP) a fin de obtener la exactitud extrema que se necesita para medir estas señales<sup>17</sup>. En esta técnica, el tiempo entre las exploraciones TO y FRO se mide entre puntos del flanco ascendente de cada pulso detectado, en el cual la pendiente alcanza un valor seleccionado. Las trayectorias

múltiples, siendo señales demoradas, no distorsionarán en forma significativa la primera parte de la forma del pulso detectado.

#### 53.4.3 Exactitud de guiage

Para el aterrizaje automático, la ICAO exige que los MLS tengan exactitudes de guiage lateral de  $\pm 6$  m (20 pies) y vertical de  $\pm 0.6$  m (2 pies). Estos son valores  $2\sigma$  y se definen para la zona de toque de tierra de la pista. En pistas largas de 3 000 a 3 500 m, el MLS ha satisfecho estas exactitudes mediante anchos de haz de las antenas de acimut y de elevación de  $1^\circ$ , con pendientes de la trayectoria de descenso (planeo), en la prueba, tan bajas como  $1^\circ 16'$ . En pistas más cortas y donde no se requiere la capacidad total de aterrizaje automático, se pueden relajar los anchos de haz de las antenas. Por ejemplo, en un sistema de tierra MLS desarrollado para aeropuertos de pista corta de pequeñas comunidades se utilizan anchos del haz de  $3^\circ$  de acimut y  $2^\circ$  de elevación, pero aun así se han demostrado a la ICAO exactitudes de guiage hasta un ángulo estándar de la trayectoria de descenso (planeo) de  $3^\circ$ .

#### 53.4.4 Equipo radiotelemétrico de precisión (PDME)

Se utiliza el PDME para definir la distancia al toque de tierra en aterrizajes MLS. Aunque no es parte de la mayoría de los formatos del MLS con el sistema básico (Az-E1) (acimut-elevación), se espera que se instale en casi todos los sitios. Originalmente se concibió como un sistema de banda C, al igual que el guiage angular MLS, pero después se demostró que los requisitos de PDME se podían satisfacer mediante el uso de un sistema de banda L, que era compatible en gran medida con el DME en ruta, descrito en la sección 52.2.

Los requisitos de funcionamiento en cuanto a distancia de los PDME actuales son:

- Aeronaves de ala fija:  $\pm 30$  m (100 pies),  $2\sigma$
- Helicópteros:  $\pm 12$  m (40 pies),  $2\sigma$



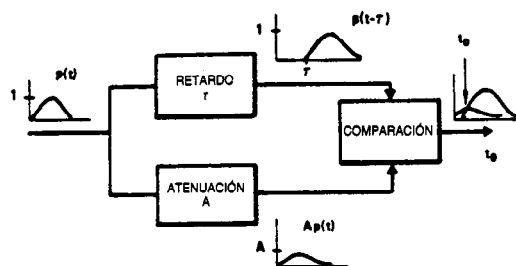


Fig. 53-8. Procesamiento de señales de retardo y comparación. Tomado de Kelly y LaBerge<sup>18</sup>, con permiso.

Estas especificaciones se deben cumplir en presencia de reflexiones intensas de trayectoria múltiple en el plano vertical, inducidas por reflexiones de tierra, y lateralmente, tales como las de los hangares de aeropuertos. En la referencia 18 se analizan varias de las técnicas de que se dispone para lograr el funcionamiento del PDME. Las velocidades de interrogación propuestas son elevadas (40 por segundo) en comparación con la dinámica de la aeronave. Esto permite realizar un filtrado de datos considerable en la nave para reducir las perturbaciones ocasionadas por el efecto de trayectoria múltiple de período largo.

Para retardos de trayectoria múltiple breves comparados con las longitudes de pulso del DME, el efecto es el de distorsionar los flancos ascendentes de los pulsos y producir variaciones grandes de un pulso a otro en las amplitudes recibidas. Una técnica que se aplica para normalizar las mediciones de los tiempos de llegada en estas condiciones es el proceso de retardo y comparación que se ilustra en la figura 53-8. El detector de pulsos no emplea un umbral fijo. Cada pulso se retarda cierta cantidad  $\tau$  y después se compara con un umbral formado por una versión atenuada (pero no retardada) de la entrada. Como se demuestra en la referencia 18, el tiempo de cruce por el umbral  $t_0$  es, en el primer orden, independiente de la amplitud de la señal de entrada y del tiempo de ascenso. El retardo  $\tau$  y la atenuación  $A$  adecuados son funciones de la longitud del pulso y de la razón de señal sobre ruido.

El esquema PDME propuesto a la ICAO para su incorporación en la especificación del MLS es una técnica bimodal de dos pulsos<sup>19</sup>. Cuando la aeronave se encuentra a más de 7 nmi del sitio de aterrizaje, interroga al transpondedor PDME de tierra con pares de pulsos espaciados 12  $\mu$ s, y éste responde con pares de pulsos espaciados 12  $\mu$ s en el modo de canal X DME estándar de la ICAO. Dentro de las 7 nmi, la aeronave conmuta a interrogaciones de alta velocidad ( $\approx 40$  por segundo) empleando pares de pulsos espaciados 18  $\mu$ s. La radiobaliza responde usando asimismo pares de pulsos espaciados 18  $\mu$ s. Las respuestas que se envían a las aeronaves cercanas y distantes se intercalan mediante el uso de los espaciamientos de pulsos adecuados. En el modo cercano, los receptores de aire y tierra utilizan el procesamiento de retardo y comparación con una atenuación ( $A$  en la fig. 53-8) de 17 dB. También se utiliza un pulso con forma aproximada de coseno-coseno<sup>2</sup>, con tiempo de ascenso de 800 ns. La

prueba estática del prototipo de PDME MLS ha demostrado errores de alcance aproximados de 5.5 m (18 pies) rms<sup>18</sup>.

#### 53.4.5 Equipo terrestre MLS

Las antenas de acimut y elevación MLS básicas se ubicarán en forma semejante a las antenas localizadora y de pendiente de la trayectoria de descenso (planeo) del ILS ordinario. La antena agregada para obtener datos precisos de elevación de enderezamiento se encontrará aproximadamente a 100 m a un costado de la pista y a 900 m del umbral. La antena de guiaje de contraacimut estará de 100 a 200 m frente al umbral de la pista, en la línea central.

El diseño específico del equipo de tierra para satisfacer los requisitos de funcionamiento del MLS varía en todos los países activos en el programa. En Estados Unidos se están produciendo dos clases de sistemas básicos de acimut y elevación (Az-El). El equipo para pistas extensas de aeropuertos importantes utiliza antenas con ancho de haz de 1°, donde cada una es un montaje en fase de 116 elementos. La electrónica de guiaje y control está alojada en un compartimiento separado con un equipo totalmente redundante. Las pistas cortas son atendidas por antenas con ancho de haz menor de 2° en elevación y 3° en acimut. En este caso se aplica un concepto de paquetes integrados; cada antena y sus componentes electrónicos importantes se combinan en una sola unidad. Además, la antena DME de precisión y los componentes electrónicos del transpondedor están integrados con el conjunto de guiaje de acimut. Para ambas versiones del equipo, actualmente se utiliza un transmisor de 20 W de potencia. Pruebas realizadas demuestran que en uso operacional deben bastar 10 W.

La figura 53-9 ilustra los dos elementos de tierra desarrollados para las fuerzas militares de Estados Unidos dentro del programa del sistema táctico conjunto de aterrizaje por microondas (JTMLS). Una unidad alberga los componentes electrónicos y antena exploradora de elevación, y la otra, el equipo y antenas de acimut y PDME. Éstos tienen un peso combinado de alrededor de 200 kg y están diseñados para dividirse en unidades adecuadas para su transporte manual. Unas baterías internas permiten su operación autónoma hasta dos horas. La antena de guiaje de elevación (El) de 2° emplea 40 elementos con defasadores a diodo para la exploración y mide unos 2 m de longitud, mientras que el montaje en fase de guiaje de acimut (Az) de 3° tiene 36 elementos en una amplitud de unos 1.25 m. El PDME, con su antena de 1.3 m de altura, se incorpora en la unidad de guiaje de acimut (Az).

#### 53.4.6 Equipo de a bordo

Por contrato con la FAA, la Bendix Corporation ha desarrollado dos tipos de equipo de avión MLS para su prueba en Estados Unidos. Algunos ejemplos del primer tipo, desarrollados según las especificaciones para líneas aéreas (ARINC), han estado en uso desde 1977.

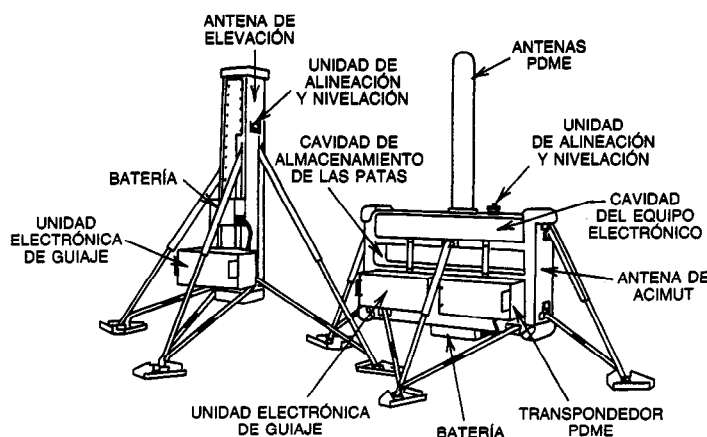


Fig. 53-9. Equipo de tierra del sistema táctico conjunto de aterrizaje por microondas (JTMLS). PDME = equipo radiotelemétrico de precisión. Tomado de Cox y Shirley<sup>15</sup>, con permiso de IEEE.

La versión más reciente (el receptor de fase III) pesa alrededor de 6 kg y está diseñado para operar junto con una unidad de a bordo DME de configuración ARINC estándar. Se ha creado un segundo tipo de unidad de a bordo para su uso final en aeronaves pequeñas. Ésta opera exclusivamente con datos de ángulos y pesa unos 3.2 kg. El RTCA ha establecido normas de funcionamiento para el MLS de a bordo<sup>20</sup>.

### 53.5 SISTEMAS DE ATERRIZAJE MILITARES PARA FINES ESPECIALES

Los servicios armados han producido diferentes sistemas para solucionar sus problemas específicos en el aterrizaje de helicópteros y aviones. En uso táctico, se espera que un sistema de aterrizaje proporcione guiage exacto en todas las condiciones meteorológicas en un área de aterrizaje relativamente pequeña y subóptima. En operaciones navales, esta "área" quizá sea la cubierta de un portaaviones o un barco con una plataforma para el aterrizaje de helicópteros. Los componentes de tierra de un sistema de aterrizaje táctico deben ser compactos, fácilmente transportables y ensamblables y susceptibles de operar en un par de horas o hasta minutos después de llegar a un nuevo sitio de ubicación. Por tanto, todos los diversos sistemas desarrollados operan en frecuencias de microondas, de manera que los haces de vigilancia y exploración estrechos que se requieren para un guiage exacto se puedan seguir obteniendo con antenas compactas.

El plan de acción de los servicios armados de Estados Unidos es que el formato de señales del sistema de aterrizaje por microondas (MLS) de la ICAO llegue a utilizarse finalmente para cumplir todos los requisitos de aterrizaje táctico<sup>1</sup>. Por tanto, el equipo de a bordo común servirá para efectuar aterrizajes en áreas tácticas y también en aeropuertos militares y civiles importantes. Para este fin, se han creado prototipos de equipo de tierra para el sistema táctico conjunto de aterrizaje por

microondas (JTMLS). Sin embargo, los diversos sistemas de aterrizaje de uso especial que ahora existen, continuarán en operación por algún tiempo.

#### 53.5.1 Sistemas de aterrizaje AIL

La división AIL de la Eaton Corporation de Farmingdale, Nueva York, ha desarrollado una familia de sistemas de aterrizaje que utilizan formatos de señales principalmente para usuarios militares, pero también con algunos usos comerciales limitados. Todos estos sistemas operan en el intervalo de 15.4 a 15.7 GHz. El guiage está provisto por pares de haces en abanico, estrechos, codificados en pulsos, que se proyectan ortogonalmente para cubrir la región de interés mediante el uso de antenas de exploración mecánica.

La figura 53-10 ilustra el principio de guiage de elevación. El haz en abanico, estrecho en elevación, transmite pares de pulsos de identidad, con un espaciamiento entre pares que varía linealmente con la elevación del haz, desde 140  $\mu$ s a 20° hasta 60  $\mu$ s a 0° en este ejemplo. La aeronave determina su ángulo de elevación midiendo el espaciamiento promedio entre pares de pulsos en su secuencia de pulsos observada. Puede emplearse un procedimiento de ajuste de la forma del haz para obtener mayor exactitud en la medición a ángulos de elevación bajos<sup>21</sup>. Para el guiage de acimut mediante el haz de barrido horizontal, el principio es el mismo, salvo que se emplea un espaciamiento fijo diferente para los "pares de identidad". Las señales de guiage de elevación y acimut se multicanalizan en tiempo sobre un canal de frecuencia común, con ritmo de actualización de 5 exploraciones por segundo. Con algunas configuraciones la aeronave puede interrogar a la estación de tierra sobre la misma frecuencia, dentro de este formato multicanalizado en tiempo, para obtener información de distancia a partir del retardo de viaje de ida y vuelta de la respuesta.

Todos estos métodos de haz explorador se derivan del sistema FLARESCAN, desarrollado por AIL

para la FAA de Estados Unidos en 1962. Ésta fue la primera demostración práctica del concepto de haz de exploración por microondas para el guiage de ángulo de elevación bajo.

**Exploración C.** Entregada a la marina militar de Estados Unidos en 1966, esta versión del sistema de aterrizaje AIL proporciona guiage a aeronaves que hacen aterrizajes en portaaviones. El sistema de guiage de a bordo del barco (designado AN/SPN-41) utiliza un haz de acimut de  $2^\circ$  y un haz de elevación de  $1.3^\circ$  para explorar un volumen de  $\pm 20^\circ$  lateralmente y de  $0$  a  $20^\circ$  en la vertical en la dirección de aproximación. La antena de elevación se estabiliza mecánicamente contra el balanceo y cabeceo del barco. Con 2 kW de potencia pico, el alcance utilizable es de unas 20 nmi. El equipo de avión presenta un visor del indicador de intersección para el piloto similar a la del equipo ILS ordinario. El ángulo correcto de pendiente de la trayectoria de descenso (planeo) se preestablece en el equipo de guiage de elevación de cada aeronave en un valor adecuado para ese tipo de nave. La exactitud es  $\pm 0.1^\circ$  en elevación y  $\pm 0.2^\circ$  en acimut. La exploración C se utiliza en forma individual, en cuyo caso el piloto cambia al guiage visual en los últimos 200 pies de descenso, o bien junto con el sistema AN/SPN-42 (que se describe más adelante) en el caso de un sistema de aterrizaje completamente automático. El AN/TRN-28 es una versión de exploración C montada sobre un camión de carga que utilizan los cuerpos de la armada y la marina de Estados Unidos para el guiage de aeronaves en algunas bases costeras.

**Sistema de aterrizaje de los transbordadores espaciales (MSBLS).** La versión de aterrizaje mediante haz de exploración por microondas (MSBLS) del sistema del AIL está siendo utilizada por la NASA para el guiage de aproximación y aterrizaje totalmente automático del vehículo orbitador transbordador espacial<sup>22</sup>. Después de interceptar inicialmente señales MSBLS a unos 14 000 pies de altitud, el transbordador recibe guiage durante los 90 s de vuelo restantes. Para el MSBLS, la cobertura de exploración de acimut es de  $\pm 20^\circ$  y la cobertura de elevación se incrementa hasta  $30^\circ$  para dar cabida al ángulo de pendiente de la trayectoria de descenso de  $21$  a  $24^\circ$  del transbordador. Las exactitudes de acimut y elevación son de  $0.05^\circ$  y  $0.03^\circ$  ( $1\sigma$ ), respectivamente, con exactitud de alcance integral (DME) de menos de 30 m. Para mejorar la confiabilidad del sistema, se incluye una doble capacidad del hardware, junto con la monitorización de campo continua de las señales transmitidas.

Hasta estos momentos se han instalado cinco estaciones de tierra MSBLS para dar servicio a los sitios de aterrizaje civiles y militares planeados del transbordador espacial.

**TILS (Sistema táctico de aterrizaje por instrumentos).** Ésta es una versión fácilmente transportable del sistema AIL que está siendo desplegada por las fuerzas aéreas de Suecia y Finlandia para proporcionar guiage de aterrizaje a sus aeronaves SAAB Viggen y Draken

en sitios de aterrizaje remotos y dispersos. Se han entregado más de 50 estaciones de tierra y 200 conjuntos de equipo de a bordo.

**Exploración A (AN/TRQ-33).** Ésta es una versión del sistema básico creado para el ejército de Estados Unidos, principalmente con fines de cobertura de acimut amplio de ángulo pronunciado para el guiage de helicópteros. Ofrece múltiples trayectorias de aterrizaje, señales de advertencia de obstáculos en tierra y DME de precisión.

**Exploración Co.** Ésta es una versión comercial del sistema AIL destinada a aeronaves de despegue y aterrizaje en corta distancia y vertical (STOL, VTOL). Se ha probado en forma intensa en Canadá en una línea aérea que da servicio a centros urbanos mediante pequeños campos de aviación de operación limitada ubicados en las ciudades; se han obtenido exactitudes de guiage equivalentes a la categoría I de ILS.

### 53.5.2 AN/SPN-42

Como parte del sistema de aterrizaje en portaaviones en todas las condiciones meteorológicas (AWCLS) de la armada de Estados Unidos, el sistema de control de aterrizaje AN/SPN-42 ha sido producido por la Bell Aerospace Company para ofrecer el aterrizaje automático en portaaviones. El SPN-42 es un radar de precisión automatizado para aproximación a bordo de barcos. La aeronave que llega es perseguida por un radar de banda Ka (32 GHz) montado sobre la superestructura "de isla" del barco. La posición y velocidad de la aeronave se comparan con las señales de seguimiento y gobierno para las correcciones deseadas, que se obtienen en una computadora a bordo del barco. El sistema AN/SPN-42 se vale de una computadora digital. En su sistema predecesor, el AN/SPN-10, y su equivalente con base en tierra, el AN/GSN-5<sup>23</sup>, se realizaban cálculos analógicos. Para el cálculo se requiere el conocimiento de los coeficientes dinámicos de la aeronave. Por tanto, antes del aterrizaje deben seleccionarse los valores correctos para cada aeronave específica. Las instrucciones de gobierno se envían a la aeronave mediante un enlace de datos de UHF de la marina y se visualizan para el piloto a través del mismo instrumento indicador de intersección que se utiliza en el guiage AN/SPN-41. En aeronaves equipadas, el acoplamiento con el piloto automático hace posible un aterrizaje totalmente automático.

La transferencia del guiage del SPN-41 al SPN-42 se realiza a una distancia aproximada de 4 nmi. Después de esa transferencia continúan las mediciones angulares del SPN-41, a fin de monitorizar el funcionamiento del SPN-42. Las señales de guiage que se transmiten son compensadas por la computadora a fin de considerar los efectos de balanceo, cabeceo y enmienda avante del barco. En los últimos segundos del guiage, esto requiere algoritmos para predecir la posición y velocidad de desplazamiento de la cubierta. La trayectoria de la aeronave se corrige a fin de minimizar las velocidades verticales relativas al tocar cubierta.

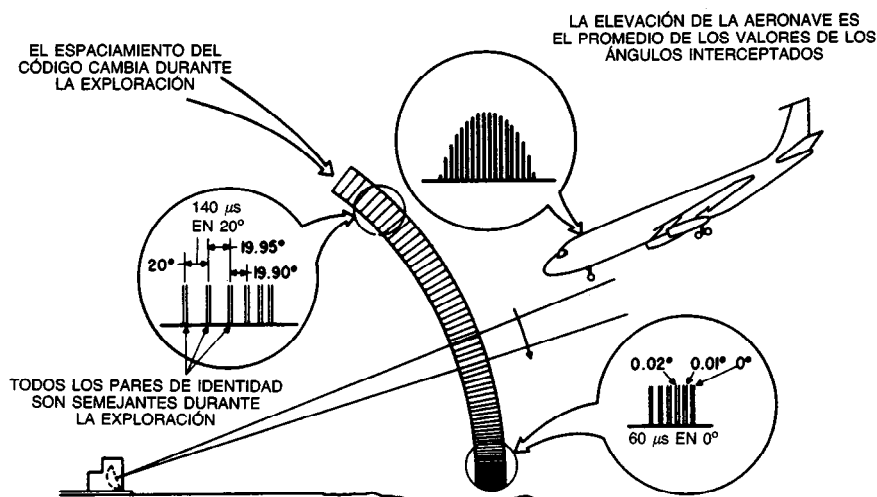


Fig. 53-10. Medición de ángulos AIL. Cortesía de AIL, división de Eaton Corporation, Farmingdale, NY.

### 53.5.3 Equipo de guiage digital de aeronaves por microondas (MADGE)

La división MEL de Philips Electronic and Associated Industries, Ltd, ha desarrollado en Inglaterra varias versiones de este sistema, creado originalmente para satisfacer un requisito militar de la OTAN (1971): un sistema auxiliar táctico portátil para la aproximación. El MADGE también se ha autorizado para uso civil en helicópteros que maniobran en las plataformas petroleras del Mar del Norte<sup>24</sup>.

El MADGE opera en varios canales de frecuencia en la banda de 5 a 5.25 GHz. El equipo de tierra o a bordo del barco no radia hasta que es interrogado por una aeronave usuaria. Al recibir una interrogación, la estación de tierra mide la posición angular de la aeronave, empleando dos montajes de antenas de interferómetro fijas ortogonales (Az-El). Por tanto, son posibles dos modos de operación.

**Modo "derivado en aire".** En este modo, los datos sobre ángulos se retransmiten a la aeronave. A partir de éstos y una estimación de la distancia que se obtiene midiendo el retardo de viaje de ida y vuelta del transpondedor, el equipo del avión puede calcular las correcciones de gobierno necesarias para llevar la aeronave al acimut y la pendiente de trayectoria de descenso (planeo) deseados seleccionados por el piloto.

**Modo "controlado desde tierra".** Se aplica cuando se emplean procedimientos de aterrizaje de trayectoria fija, particularmente para aterrizajes en plataformas petroleras de alta mar. Las desviaciones se calculan en la estación de tierra y se transmiten a la aeronave en forma de señales de desviación fuera de ruta. Este modo es especialmente adecuado en trayectorias de aterrizaje desviadas que no intersecan el "origen" representado por la ubicación física de las antenas de guiage.

La interrogación de aire a tierra en el modo derivado en aire normal, consta de un pulso marcador de 3  $\mu$ s, seguido de 25 bits de datos, entre ellos los códigos de dirección (en registros) de la aeronave y de la estación de tierra. Éstos son transmitidos a razón de 1 MHz empleando modulación de amplitud de pulsos con unos 250 W de potencia radiada. Esto se repite a razón de 50 Hz. La respuesta de la estación de tierra consta de un pulso marcador más 60 bits de datos en el mismo formato. Para el modo de control en tierra, los mensajes de datos de aire y de tierra son más largos y se utiliza una razón de interrogación de 100 Hz. Los códigos digitales de la aeronave y de tierra permiten la operación hasta de 25 sistemas de tierra muy próximos, en la misma frecuencia, aunque se dispone de 12 grupos de frecuencia distintos.

La antena de medición de acimut en tierra es un montaje de interferómetro horizontal de siete bocinas, con espaciamiento desigual en un ancho aproximado de 1.5 m. La antena de elevación utiliza interferómetros bastos y finos de cuatro y ocho elementos, respectivamente, sobre una altura aproximada de 2 m. Una tercera unidad de tierra es la radiobaliza de respuesta o transpondedor, que tiene una antena de 1 m de alto y radia pulsos con potencia pico de 150 W. El peso total de la estación de tierra es de alrededor de 100 kg.

Las dimensiones de cobertura son comúnmente de  $\pm 45^\circ$  en sentido horizontal y de  $25^\circ$  en sentido vertical desde la estación de tierra, con un alcance de operación de 15 nmi. La exactitud de guiage se da como de 0.1° rms en ángulo y de  $\pm 30$  m en distancia. El MADGE ha sido autorizado para aproximaciones de aterrizaje hasta una altura de descenso mínima de 30 m (100 pies) y 400 m de alcance visual, equivalente a las normas de la categoría II de ILS de ICAO. Ahora está en producción para equipar los portaaviones y aeronaves Harrier VTOL de la armada real británica (British Royal Navy).

### 53.5.4 Sistema de aproximación y aterrizaje de la marina en sitios remotos (MRAALS)

Este sistema ha sido desarrollado por la Singer-Kearfott Company de New Jersey bajo contrato con los marines de Estados Unidos. El MRAALS está diseñado para ofrecer a los helicópteros de los marines un recurso de aterrizaje en todas las condiciones meteorológicas, en sitios remotos, en un entorno táctico. Es un sistema de haz de exploración por microondas (de 15 GHz), con formato de señales idéntico al del sistema de exploración C de AIL (AN/SPN-41) descrito anteriormente.

Mediante el uso de antenas de exploración mecánica, la estación de tierra de una sola unidad (designada AN/TPN-30) proporciona señales de guiage sobre un volumen de  $\pm 20^\circ$  en acimut, 0 a  $20^\circ$  en elevación y 10 nmi de distancia. La exactitud es de  $\pm 0.1^\circ$  en acimut,  $\pm 0.05^\circ$  en elevación y  $\pm 30$  m en distancia mediante el uso de un sistema radiotelemétrico asociado compatible con el TACAN. El peso total de la estación de tierra es de unos 50 kg.

#### REFERENCIAS BIBLIOGRÁFICAS

1. F. G. Kear, "Instrument Landing at the National Bureau of Standards", *IRE Trans. Aeronaut. Nav. Electr.* (Artículo especial sobre sistemas de aterrizaje por instrumentos.)
2. *International Standards and Practices, Aeronautical Telecommunications*, Anexo 10 de la Convención on International Civil Aviation, ICAO, Montreal, Canadá, 3.<sup>a</sup> ed.
3. W. E. Jackson, "Improvements on the Instrument Landing System", *IRE Trans. Aeronaut. Nav. Electr.*
4. "Localizer-Mark ID System", Northrop-Wilcox Corp., Kansas City, MO.
5. H. H. Butts y R. H. McFarland, "New Developments in Instrument Landing System", *IEEE Trans. Aero. Electr. Syst.* AES-2(6).
6. J. J. Battistelli, "The Conventional ILS—So What's New?", *Navigation*, Publicación de The Institute of Navigation, 21(2).
7. F. W. Iden, "Glide Slope Antenna Arrays for Use under Adverse Siting Conditions", *IRE Trans. Aeronaut. Nav. Electr.*
8. R. H. McFarland, "Application of End-Fire Arrays at Contemporary Glide-Slope Problem Sites", *IEEE Trans. Aero. Electr. Syst.* AES-17(2).
9. "Automatic Landing—Now a Matter of Course", *Interavia*.
10. *Minimum Performance Standards—Airborne ILS Receiving Equipment*, DO-131A (localizador), DO-132A (trayectoria de descenso), RTCA, Washington, DC.
11. *Federal Radio Navigation Plan*, DOT-TSC-RSPA-80-16, Department of Defense/Department of Transportation (4 vols.).
12. *Minimum Performance Standards—Airborne Low-Range Radar Altimeters*, DO-155, RTCA, Washington, DC.
13. M. I. Skolnik, *Introduction to Radar Systems*, McGraw-Hill, Nueva York.
14. H. R. Ward y colaboradores, "GCA Radars: Their History and State of Development", *IEEE Proc.* 62(6).
15. R. M. Cox y J. M. Shirey, "MLS—A New Generation Landing System is Here", presentado en IEEE Position Location and Navigation Symposium (PLANS 80), Atlantic City, NJ.
16. R. J. Kelly, H. W. Redlien y J. L. Shagena, "Landing Aircraft under Poor Conditions", *IEEE Spectr.*
17. R. J. Kelly y J. T. Skudrna, "Joint Tactical Microwave Landing System (JTMLS) Airborne Signal Processing", presentado en la 4<sup>a</sup> AIAA/IEEE Digital Avionics Systems Conference, St. Louis, MO.
18. R. J. Kelly y E. F. LaBerge, "Guidance Accuracy Considerations for the Microwave Landing System Precision DME", *Navigation*, Publicación de The Institute of Navigation, 27(1).
19. Final Report, ICAO All-Weather Operations Panel, Working Group M, ICAO, París.
20. *Minimum Operational Performance Standards—MLS Airborne Receiving Equipment*, DO-177, RTCA, Washington, DC.
21. G. Blazek y A. Charych, "Beam-Fit Decoding Technique for Increasing Elevation Measurement Accuracy from Scanning Beam Systems at Low Elevation Angles", AIAA Guidance and Controls Conference, Colorado Springs, CO, AIAA 79-1710.
22. G. Blazek y L. M. Carrier, "Microwave Landing System Selected for Space Shuttle Program", *ICAO Bull.*
23. F. D. Powell, "An Automatic Landing System", *IRE Trans. Aeronaut. Nav. Electr.*
24. D. A. Brown, "MADGE Certification Expected in March", *Aviat. Week Space Technol.*

## CAPÍTULO 54

# Navegación por satélite

**John H. Painter**

Texas A & M University  
College Station, Texas

- |   |  |
|---|--|
| <ul style="list-style-type: none"><li><b>54.1 Principios de la radionavegación por satélite</b></li><li><b>54.2 Descripción de la órbita de un satélite (ecuación de Kepler)</b></li><li><b>54.3 Coordenadas del usuario y geodesia</b></li><li><b>54.4 Adquisición de datos de navegación</b></li><li><b>54.5 Efectos ionosféricos</b></li><li><b>54.6 TRANSIT, sistema Doppler de baja dinámica</b><ul style="list-style-type: none"><li>54.6.1 Señales de los TRANSIT y estructura del sistema</li><li>54.6.2 Técnica de localización Doppler de TRANSIT</li><li>54.6.3 Rendimiento y características de error del TRANSIT</li></ul></li></ul> | <ul style="list-style-type: none"><li><b>54.7 GPS NAVSTAR, sistema telemétrico de alta dinámica</b><ul style="list-style-type: none"><li>54.7.1 Señales y estructura del sistema GPS</li><li>54.7.2 Técnica de telemetría por pseudoruido (PN) del GPS</li><li>54.7.3 Técnicas de receptores en coherencia e incoherencia de fase</li><li>54.7.4 Procesamiento de la navegación: linealización y dilución geométrica de la previsión (GDOP)</li><li>54.7.5 Filtrado de Kalman: estándar y adaptativo</li><li>54.7.6 Filtrado de Kalman extendido para el GPS</li><li>54.7.7 Rendimiento del GPS y clasificación de errores</li></ul></li></ul> |
|---|--|

#### 54.1 PRINCIPIOS DE LA RADIONAVEGACIÓN POR SATÉLITE

Los conceptos matemáticos básicos subyacentes a la radionavegación por satélite no son diferentes de los de cualquier otro tipo de navegación en el que se utilicen mediciones independientes para determinar la posición. Primero se elige un sistema de coordenadas por medio del cual se describe la posición del navegante. Para los fines de este capítulo, se puede considerar un sistema cartesiano derecho con las componentes  $x$ ,  $y$  y  $z$  y origen en el centro de la Tierra. Este sistema se denomina centrado en la Tierra o geocéntrico. Es una práctica usual considerar que el eje  $z$  tiene su dirección positiva hacia el polo norte. De esta forma, los ejes  $x$  e  $y$  yacen en el plano del ecuador. Este marco o sistema de referencia coordinado se denomina ecuatorial, a diferencia del sistema coordinado eclíptico que se utiliza en la astronomía y la navegación astronómica o interplanetaria.

Si el eje  $x$  se proyecta desde el centro de la Tierra hasta un punto fijo en el espacio situado a una distancia infinita, tal como el equinoccio vernal en la esfera celeste, el sistema coordinado no gira con la Tierra. Si el eje  $x$  se proyecta pasando por un punto sobre la superficie terrestre, tal como el punto de latitud cero y longitud cero, el sistema se denomina geocéntrico fijo en la Tierra (ECEF).

En este capítulo se considera uno u otro tipo de sistema coordinado cartesiano geocéntrico, según las necesidades.

Ahora se definirá la posición del navegante en el sistema  $x$  y  $z$  como una matriz de tres vectores o matriz columna de  $3 \times 1$ , como:

$$\mathbf{P} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (54-1)$$

Si el sistema coordinado de  $\mathbf{P}$  es ECEF, existen transformaciones bien conocidas para expresar  $\mathbf{P}$  en latitud, longitud y altitud, por ejemplo sobre una Tierra esférica perfecta.

Las correcciones correspondientes a la Tierra no esférica real resultan más complejas y se analizan en la sección 54.3.

En la navegación en el espacio tridimensional se hacen tres mediciones "independientes" de cantidades que se sabe se relacionan funcionalmente con la posición del navegante. Dado un conocimiento de las relaciones matemáticas funcionales que conectan las componentes de la posición y las mediciones, la posición se "despeja" a partir de las mediciones. Esta solución existe siempre que las mediciones sean adecuadas para el problema de navegación. Por ejemplo, las mediciones de la distancia de un navegante a tres puntos conocidos producen un número finito de soluciones<sup>1</sup> siempre que los tres puntos conocidos no sean colineales. La solución ambigua es rechazada por el conocimiento *a priori* que el navegante tiene de su posición.

Si las tres mediciones se representan por  $m_1$ ,  $m_2$  y  $m_3$ , en un problema de navegación bien planteado existen tres funciones escalares  $f_1(\cdot)$ ,  $f_2(\cdot)$  y  $f_3(\cdot)$  tales que:

$$\begin{aligned} m_1 &= f_1(x, y, z) \\ m_2 &= f_2(x, y, z) \\ m_3 &= f_3(x, y, z) \end{aligned} \quad (54-2)$$

Si ahora las mediciones y funciones se representan por medio de tres vectores, como:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} f_1(\cdot) \\ f_2(\cdot) \\ f_3(\cdot) \end{bmatrix} \quad (54-3)$$

Entonces se tiene una notación compacta de la relación funcional como:

$$\mathbf{m} = \mathbf{f}(\mathbf{P}) \quad (54-4)$$

donde  $\mathbf{f}(\cdot)$  es una función con valor vectorial del argumento vectorial,  $\mathbf{P}$ .

El problema de navegación es: "Dados el sistema de referencia de la función  $\mathbf{f}(\cdot)$ , y las mediciones  $\mathbf{m}$ , despejar  $\mathbf{P}$ ." Si, por ejemplo, la relación entre  $\mathbf{m}$  y  $\mathbf{P}$  fuese lineal inversible, el problema sería simple. Por desgracia, la mayor parte de los problemas de navegación carecen de una función inversa para  $\mathbf{f}$  que se pueda aplicar en forma directa a fin de despejar  $\mathbf{P}$ .

Por ejemplo, en el problema de la medición de las distancias a partir de la posición del navegante respecto a los tres puntos descritos por los vectores  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ ,  $\mathbf{P}_3$ , las componentes de la función  $\mathbf{f}(\cdot)$  están dadas por:

$$f_i(\mathbf{P}) = \|\mathbf{P}_i - \mathbf{P}\| \quad i = 1, 2, 3 \quad (54-5)$$

Es decir, la función es la raíz cuadrada del producto escalar, o la norma euclidiana. Esta función es no lineal y no tiene inversa.

En casos en que la función  $\mathbf{f}(\cdot)$  no tiene inversa, suele utilizarse la llamada solución por perturbaciones, basada en la "linealización" del problema de navegación. El procedimiento se expone a continuación.

Supóngase que existe una estimación razonablemente precisa de la posición del navegante,  $\mathbf{P}_0$ . A partir de esta posición supuesta  $\mathbf{P}_0$ , calcúlese la medición "predicha" correspondiente  $\mathbf{M}_0$  empleando la función conocida  $\mathbf{f}(\cdot)$ . Esto es:

$$\mathbf{M}_0 = \mathbf{f}(\mathbf{P}_0) \quad (54-6)$$

Ahora bien,  $\mathbf{M}_0$  es el primer término de una serie vectorial de Taylor para  $\mathbf{M}$ , desarrollada alrededor del "punto"  $\mathbf{P}_0$ . Calcúlese el coeficiente del segundo término lineal de la serie, que es la matriz jacobiana  $H$ :

$$H = \left. \frac{\partial \mathbf{f}(\cdot)}{\partial \mathbf{P}} \right|_{\mathbf{P}=\mathbf{P}_0} \quad \text{una matriz de } 3 \times 3 \quad (54-7)$$

Entonces:

$$\mathbf{M} = \mathbf{M}_0 + H(\mathbf{P} - \mathbf{P}_0) \quad (54-8)$$

La ecuación 54-8 es una aproximación (lineal) de dos términos de  $\mathbf{M}$ . Si existe una solución de navegación,  $\mathbf{H}$  será inversible con inversa  $H^{-1}$ , y  $\mathbf{P}$  puede expresarse como:

$$\mathbf{P} = \mathbf{P}_0 + H^{-1}(\mathbf{M} - \mathbf{M}_0) \quad (54-9)$$

Esta solución linealizada por perturbaciones de la posición del navegante es en la actualidad una práctica común en los sistemas de navegación basados en la computadora.

En un sistema de radionavegación por satélite, es función del receptor de radio del navegante proporcionar las mediciones  $m_1$ ,  $m_2$  y  $m_3$ , que comprenden a  $\mathbf{M}$ . En el sistema TRANSIT, por ejemplo, estas mediciones son de los valores de desplazamiento Doppler de la radiofrecuencia transmitida por el satélite, debidos a la velocidad de éste. En el sistema de localización global (GPS), las mediciones son de las distancias, o alcances, desde los satélites hasta el receptor del navegante. En los sistemas GPS y TRANSIT, la computadora de navegación requiere el conocimiento de la posición del satélite en el momento (instante) de hacer la medición. Así, ambos sistemas también incorporan en la transmisión de radio los mensajes de datos del satélite al navegante. Estos mensajes contienen los elementos orbitales keplerianos a partir de los cuales es posible calcular la órbita del satélite en el procesador de navegación.

Para ser altamente exactos, los sistemas de radionavegación por satélite deben contener medios para corregir los errores inducidos por el paso de las señales de radio a través de la ionosfera. Las alturas orbitales de los sistemas TRANSIT y GPS son suficientemente grandes para que las trayectorias de las señales de radio hacia un navegante cerca de la superficie de la Tierra atraviesen una parte importante de la ionosfera. Los efectos del paso por la ionosfera en la señal de radio son más pronunciados cuanto menores son la frecuencia de la señal y el ángulo de elevación del satélite visto desde la posición del navegante. En los dos sistemas se aplican técnicas para corregir los errores de localización inducidos por la ionosfera.

Los sistemas de radionavegación como TRANSIT y GPS, en los que se emplean radiotransmisiones por satélite, se denominan pasivos en virtud de que el sistema de navegación es sólo receptor. No se requieren transmisiones del navegante, como en el equipo radiotelemétrico (DME) de navegación aérea táctica (TACAN). Estos canales de comunicación enviados desde el espacio operan exclusivamente en un modo de línea de visión (trayectoria óptica) y utilizan transmisiones de potencia relativamente baja desde los satélites. Es típica una potencia de transmisión de 20 W o menos. Por tanto, los niveles de potencia de señal disponibles para el receptor son muy bajos, del orden de -160 dBW para los sistemas TRANSIT y GPS. Con cifras de ruido en el receptor del orden de 3 dB, ra-

zones de potencia de señal disponible sobre densidad espectral de potencia del ruido del orden de 40 a 50 dBHz hacen necesario un procesamiento de señales eficiente.

En las siguientes secciones se analizan con mayor detalle algunos de los elementos de radionavegación por satélite que son comunes a todos los sistemas. Son ejemplos de ello la descripción orbital de los satélites y los efectos ionosféricos. Después de esas exposiciones, se describen con cierta profundidad los sistemas TRANSIT y GPS.

## 54.2 DESCRIPCIÓN DE LA ÓRBITA DE UN SATELITE (ECUACIÓN DE KEPLER)

No se pretende exponer de manera completa la mecánica orbital en esta sección; para ello, el lector puede consultar alguno de los textos estándares, como el de Battin<sup>2</sup>. Sin embargo, se requiere que el lector tenga ciertos conocimientos, ya que los receptores procesadores de radionavegación por satélite contemporáneos, tales como los de los sistemas TRANSIT y GPS, realizan los cálculos de las órbitas de los satélites en forma interna. Además, los mensajes de datos que se envían desde el satélite contienen los llamados elementos orbitales del satélite que transmite. Por tanto, es necesario tener un conocimiento de estos elementos y de su uso para entender el sistema mismo.

La forma de una órbita estable de satélite es una elipse cuando sólo se consideran el satélite y la Tierra. Esto es, si se considera que el satélite y la Tierra forman un "problema de dos cuerpos" de mecánica celeste, la solución de la trayectoria del satélite es una elipse con la Tierra (centro de masa) en un foco. La órbita elíptica es descrita fundamentalmente por su semieje recto, su excentricidad y el tiempo de la mayor aproximación a la Tierra. Estas tres cantidades se denominan elementos orbitales. Para relacionar esta órbita elíptica con el sistema de coordenadas en uso se requieren tres elementos más, que se toman clásicamente como los ángulos de Euler (que se definen más adelante).

En la figura 54-1 se ilustran varios de los elementos orbitales que se utilizan para describir las órbitas de satélites de radionavegación contemporáneos. Se identifican las siguientes cantidades con respecto a la figura 54-1:

$\angle AFP = V$ : anomalía real

$\angle ACQ = E$ : anomalía excéntrica

$\overline{AC} = a$ : semieje mayor

$\overline{FP} = r$ : distancia radial orbital

$0 \leq e \leq 1$ : excentricidad orbital

$P$ : posición del satélite en la órbita

$A$ : punto de perigeo

$F$ : punto focal de la elipse (posición del centro de masa de la Tierra)

$C$ : centro de la circunferencia circunscriptora

$\mu$ : parámetro gravitacional universal de la Tierra

$$n_0 = \frac{2\pi}{T} = \sqrt{\frac{\mu}{a^3}} : \text{movimiento angular medio}$$



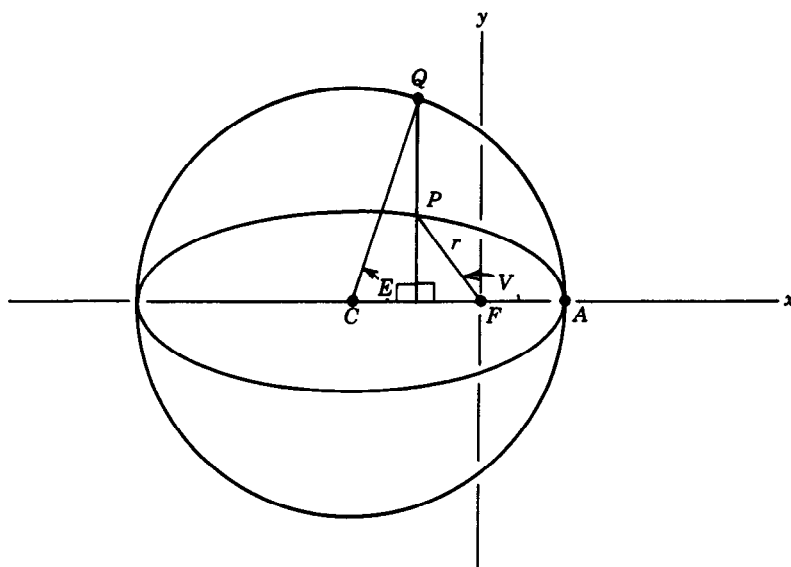


Fig. 54-1. Parámetros de órbita elíptica.

$T$ : periodo orbital  
 $t_{oe}$ : tiempo de efemérides  
 $M$ : anomalía media

La distancia radial orbital,  $r = \overline{FP}$ , está dada por:

$$r = a(1 - e \cos E) \quad (54-10)$$

Sea que  $x''$  e  $y''$  representen un sistema coordenado cartesiano local en el plano de la elipse con el eje  $x''$  que atraviesa el foco y el punto de perigeo. El eje  $y''$  es ortogonal a  $x''$  y pasa por el foco. Entonces, las coordenadas del punto  $P$  en la órbita elíptica serán:

$$\begin{aligned} x'' &= a(\cos E - e) = r \cos V \\ y'' &= a(1 - e^2)^{1/2} \sin E = r \sin V \end{aligned} \quad (54-11)$$

La anomalía excéntrica  $E$  se determina resolviendo la ecuación de Kepler, dada por:

$$n(t - t_{oe}) = M = E - e \sin E \quad (54-12)$$

En la ecuación 54-12, la cantidad  $t$  es el momento en el cual se determinará el valor de  $E$  (y de  $x''$  e  $y''$ ). Los valores de  $t$  y  $t_{oe}$  deben medirse con el mismo reloj.

Las ecuaciones 54-11 y 54-12 definen la posición orbital en función del tiempo, referida al plano orbital, al eje mayor y al tiempo de perigeo. Estas ecuaciones definen un sistema coordenado cartesiano geocéntrico derecho  $x'', y'', z''$ , para el cual  $z''$  es idénticamente cero.

Ahora existen otros dos sistemas cartesianos geocéntricos derechos de interés y en los cuales se debe definir la órbita. Ambos se conocen como ecuatoriales, ya que en ellos los ejes  $x$  e  $y$  yacen en el plano del ecuador. El eje  $z$  en ambos sistemas atraviesa el polo norte real. El primer sistema coordenado se llama

inercial (sólo para la Tierra), ya que el eje  $x$  está dirigido a un punto fijo en el firmamento. Ese punto es el primero de Aries, o equinoccio vernal. Es el punto en el cual el Sol asciende sobre el plano del ecuador. El segundo sistema coordenado recibe el nombre de fijo en tierra, ya que el eje  $x$  pasa por el meridiano de Greenwich.

En un problema de dos cuerpos con una Tierra de densidad de masa esféricamente simétrica y sin fricción de aire o vuelo alimentado por energía, la órbita elíptica es fija respecto al sistema coordenado inercial. Por tanto, si el sistema inercial se denota como  $x', y', z'$ , estas nuevas coordenadas se pueden obtener a partir de  $x'', y'', z''$  por medio de tres rotaciones sucesivas simples (de un eje) con respecto al origen (el centro de la Tierra).

Las rotaciones que se requieren se visualizan mejor con ayuda de la figura 54-2. En ésta, el plano orbital se muestra inclinado hacia el plano ecuatorial en un ángulo de inclinación  $i$ . Asimismo, el eje mayor orbital, que pasa por el punto de perigeo, se desplaza en el plano orbital desde el punto de intersección con el plano ecuatorial un ángulo  $\omega$ , llamado argumento del perigeo. El punto de intersección entre los dos planos con la coordenada  $z$  del satélite creciente se denomina cruce nodal ascendente. La intersección de los dos planos, llamada línea de nodos, se desplaza angularmente hacia el este del eje  $x'$  en el plano ecuatorial una cantidad  $\Omega$ . Este último ángulo se denomina ascensión recta del nodo ascendente. Estos ángulos son los ángulos de Euler.

A partir de esta visión de las relaciones angulares entre la órbita elíptica en  $x'', y'', z''$  y el marco inercial geocéntrico en  $x', y', z'$ , es claro que se necesitan rotaciones para transformar el primero en el segundo. El orden de rotación es  $\omega, i$  y  $\Omega$ . La transformación compuesta se expresa en forma de matriz vectorial como:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \Lambda \cdot \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} (\cos \omega \cos \Omega) - \sin \omega \cos i \sin \Omega - (\sin \omega \cos \Omega) + \cos \omega \cos i \sin \Omega - (\sin i \sin \Omega) \\ (\cos \omega \sin \Omega) + \sin \omega \cos i \cos \Omega - (\sin \omega \sin \Omega) - \cos \omega \cos i \cos \Omega - (\sin i \cos \Omega) \\ (\sin \omega \sin i) & (\cos \omega \sin i) & (\cos i) \end{bmatrix} \quad (54-13)$$

La transformación  $\Lambda$  lleva el satélite al sistema coordenado inercial geocéntrico  $x', y', z'$  en una rotación en torno al eje rotacional de la Tierra en un ángulo  $\Omega$ , que es la ascensión recta del nodo ascendente. Para llevar la posición al sistema ECEF,  $x, y, z$  donde  $x$  se proyecta a través del meridiano de Greenwich en el ecuador, se puede redefinir  $\Omega$  y utilizar la transformación  $\Lambda$ .

Sea  $\Omega$  redefinido como la longitud del nodo ascendente, medido positivo en dirección oeste desde el meridiano de Greenwich. Ahora, el ángulo  $\Omega$  es variante en el tiempo, debido a la rotación de la Tierra y a la precesión del plano orbital en el espacio inercial. Por tanto se define:

$$\Omega = (\Omega_0 - \dot{\Omega}_e \cdot t_{oe}) + (\dot{\Omega} - \dot{\Omega}_e)(t - t_{oe}) \quad (54-14)$$

En la ecuación 54-14, la primera expresión entre paréntesis da  $\Omega_0$ , en longitud del nodo ascendente en  $t = 0$ , corregida por la velocidad de rotación hacia el este de la tierra  $\dot{\Omega}_e$ , que actúa hasta que  $t = t_{oe}$ . La segunda expresión entre paréntesis corrige  $\Omega$  al tiempo  $t$  para la rotación de la Tierra y la precesión del plano orbital a la velocidad  $\dot{\Omega}$ , positiva hacia el oeste. El parámetro  $t_{oe}$  se denomina tiempo de era y se alinea con el tiempo del paso por perigeo.

Debido a que un problema de dos cuerpos no es lo suficientemente exacto, ni el campo gravitatorio de la Tierra es esféricamente simétrico y, por otras razones, los parámetros de la órbita no son estables. Por tanto, se necesita una gran cantidad de correcciones de tiempo real para obtener cálculos de órbitas exactos. En particular, el movimiento medio  $n$ , la inclinación  $i$ , la ascensión recta del nodo ascendente  $\Omega$  y la anomalía media  $M$  se especifican en  $t_{oe}$  y se corrigen en tiempos subsiguientes empleando diversos factores de corrección. Asimismo, el "argumento de latitud",  $\phi = v + \omega$ , y el radio de la órbita  $r$  y la inclinación  $i$  del satélite, se corrigen en cuanto al segundo armónico de zona del campo gravitatorio.

El orden de los cálculos, con las correcciones, se indica en la tabla 54-1.

### 54.3 COORDENADAS DEL USUARIO Y GEODESIA

Aunque los sistemas de navegación por satélite como el TRANSIT y GPS operan en un sistema coordenado cartesiano (ECEF) ortogonal derecho, el usuario en realidad no lo hace así. El sistema de coordenadas del usuario más común es el de latitud, longitud y altitud (LLA), sistema que es más o menos diferencialmente ortogonal dentro de una vecindad local del usuario.

Los usos principales del sistema LLA están en la localización del usuario sobre una carta de navegación o con respecto a un punto conocido. Los usuarios marítimos por lo general no se preocupan por la altitud (o altura). Los usuarios aeronáuticos suelen estar interesados en distancias y rumbos de su posición respecto a un punto conocido.

Existen dos factores que complican la navegación por satélite en cuanto a la localización de la posición del usuario en el sistema LLA en una carta de navegación. El primer problema es que la Tierra no es una esfera. El modelo terrestre aceptado universalmente es un elipsoide de revolución, simétrico con respecto al eje de rotación de la Tierra. Este elipsoide es una esfera "achatada" en las regiones polares. Se define un coeficiente de achatamiento  $f$ , relacionado con la excentricidad  $e$  de la elipse generatriz por medio de:

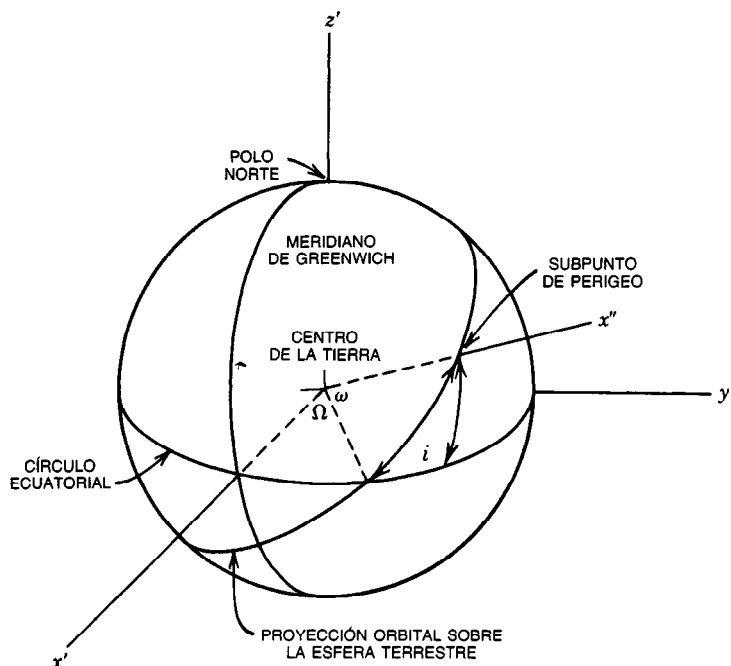
$$e^2 = 2f - f^2 \quad (54-15)$$

El modelo actual universal de la Tierra es el sistema geodésico mundial de 1972 (WGS-72). Éste es un elipsoide que produce un "mejor ajuste" al geoide de la Tierra real. El coeficiente de achatamiento del WGS-72 es:

$$f = 1/298.26 \quad (54-16)$$

Una consecuencia del achatamiento del modelo esferoide de la Tierra es que ahora la altitud y la latitud se deben medir con respecto al elipsoide WGS-72 y no con respecto a un esferoide. En teoría, la longitud sigue siendo la misma. Sin embargo, la latitud y la altitud se determinan ahora como cantidades "geodésicas".

En el caso de un esferoide perfecto con distribución de la masa esféricamente simétrica, el radio vector al centro de la Tierra, el vector normal a la superficie y la vertical gravitacional están en la misma dirección, independientemente de la posición sobre la superficie. En el caso de un elipsoide de revolución, estas tres direcciones no necesariamente coinciden, salvo en los polos y en el ecuador. Por tanto, la latitud en una posición dada se mide entre la intersección en línea de los planos ecuatorial y meridional y un vector de posición normal a la superficie del elipsoide cuya longitud es el radio de curvatura en la superficie. El ángulo subtendido de este modo se denomina latitud geodésica  $\lambda$ . El vector de posición normal a la superficie no tiene el origen en el centro del elipsoide. Esta latitud geodésica no es la misma que la latitud de un esferoide ni es igual a la latitud astronómica, que se basa en la vertical gravitacional local.



**Fig. 54-2.** Conversión de coordenadas de órbitas.  $i$  = ángulo de inclinación;  $\omega$  = argumento del perigeo;  $\Omega$  = ascensión recta del nodo ascendente.  $i$ ,  $\omega$  y  $\Omega$  = ángulos de Euler.

Existe una relación de forma cerrada entre el vector de posición ECEF  $[x, y, z]^T$  y el vector LLA geodésico del usuario  $[\lambda, \phi, h]^T$ . Estas relaciones se representan en la tabla 54-2.

La longitud  $\lambda$ , positiva en dirección oeste, se puede despejar directamente del modelo modo:

$$\lambda = \arctan \left( \frac{-y}{x} \right) \quad (54-17)$$

Sin embargo, la latitud  $\phi$  y la altitud  $h$  requieren una solución implícita, análoga a la de la ecuación de Kepler (ec. 54-12). En la referencia 1 se describe un método específico.

El segundo problema en el establecimiento de la posición del usuario sobre una carta de navegación es el del punto de referencia. Durante muchos años, en diversas regiones del mundo se han utilizado elipsoides particulares ajustados a la región local. Así, existen muchos sistemas de puntos de referencia localizados que se utilizan en la elaboración de mapas. En Estados Unidos, se utiliza el punto de referencia de Norteamérica. En Europa, se emplea el punto europeo. En Japón se utiliza el punto de Tokio, etc. En la referencia 3 se presenta una lista parcial.

Existe un conjunto estándar de ecuaciones para trasladar posiciones ECEF de algún otro punto al del WGS-72. Éstas son las denominadas fórmulas de Molodensky<sup>4</sup>, que se pueden utilizar para definir posiciones respecto al WGS-72 a partir de mapas de estudio locales o para localizar posiciones en el WGS-72 sobre mapas locales.

Parecería que con el advenimiento de la navegación por satélite, la geodesia de todo el mundo podría gravitar pronto hacia el punto de referencia del WGS-72 ya sea en su forma de 1972 o en formas posteriores, tal como la de 1982. En realidad, como la exactitud de la localización por satélite tiende a la escala submétrica, la geodesia más exacta del mundo pronto se determinará a partir de satélites.

#### 54.4 ADQUISICIÓN DE DATOS DE NAVEGACIÓN

En las secciones anteriores se ha observado que se requiere una gran cantidad de datos subsidiarios en un receptor de navegación por satélite. La mayor parte de estos datos se refieren a las órbitas de los diversos satélites y los cálculos consecuentes. Muchos datos consisten en parámetros cuyos valores cambian continuamente. Debido a la exactitud que se requiere al determinar posiciones de satélites, es necesario conocer los parámetros orbitales con bastante exactitud cada vez que se calcule una posición del usuario.

La forma más simple en que un usuario puede obtener estos datos variables en el tiempo es desde el satélite mismo. Como los satélites son seguidos por sus estaciones de control en la Tierra, las órbitas se calculan en forma continua con gran exactitud. De este modo, se predicen con mucha anticipación tablas de parámetros y correcciones orbitales keplerianas, que se transmiten a los satélites para su almacenamiento. En un momento dado, un satélite transmite en modo de

emisión la tabla de datos para ese satélite en particular en ese momento específico. Cualquier usuario que recibe señales de un satélite determinado también recibe los elementos orbitales actuales de ese satélite como parte de un mensaje de "enlace de datos". La señal de enlace de datos suele incorporarse a la señal que se utiliza para hacer la medición de navegación misma. Las mecanizaciones exactas de las señales de enlace de datos y de medición de los sistemas TRANSIT y GPS se describen con detalle en una sección posterior.

Es importante comprender que las técnicas de navegación por satélite que se detallan aquí y que se incorporan en los sistemas TRANSIT y GPS son completamente pasivas por parte del usuario. Esto es, el usuario no transmite al satélite, como en un sistema de navegación con radar transpondedor. El usuario sólo necesita recibir señales del satélite a fin de acumular mediciones y datos auxiliares suficientes para resolver su propia posición. El modo pasivo sólo de recepción de la radionavegación tiene grandes ventajas en términos de tamaño, costo y confiabilidad del equipo del usuario. La desventaja respecto a los métodos activos de transmisión y recepción está en la cantidad de cálculos que deben realizarse en el receptor del usuario. El TRANSIT y el GPS son esencialmente sistemas basados en la computadora del usuario.

#### 54.5 EFECTOS IONOSFÉRICOS

La ionosfera es la interficie entre la atmósfera y el espacio exterior. Es una región de gases atmosféricos ionizados y electrones libres que no se han recombinado con los iones atmosféricos de donde provienen. Históricamente, la ionosfera se describía como formada por "estratos" o capas que se creía causaban la "reflexión" de ondas de radio procedentes de la Tierra. Ahora se sabe que la ionosfera es descrita por una densidad de electrones por centímetro cúbico que comienza abruptamente en la vecindad de los 50 km de altitud, tiene una densidad máxima en la cercanía de los 250 a 450 km y disminuye más gradualmente hasta cero alrededor de los 800 a 3 000 km.

Una onda de radio que asciende a la ionosfera está sujeta a efectos nocivos cuya magnitud es inversamente proporcional al cuadrado de la frecuencia de transmisión. En el caso de frecuencias adecuadamente bajas, tales como en el intervalo de 15 a 1 500 kHz, durante el día (cuando la densidad iónica es máxima) las señales están sujetas a absorción en la región ionosférica más baja, la región D. Las señales de más alta frecuencia que pasan a través de la región D están sujetas a una refracción uniforme, o flexión gradual de sus trayectorias de rayos, en las regiones superiores, llamadas E, F<sub>1</sub> y F<sub>2</sub>, en orden de altitud creciente. Si la densidad de electrones es lo suficientemente grande con respecto a la frecuencia de las señales, la onda se puede refractar lo suficiente para volver a Tierra. Este comportamiento se conoce como efecto de salto.

La presencia de la ionosfera y su densidad se deben al bombardeo solar y a la radiación de las moléculas atmosféricas. Así, la ionosfera es más densa en torno

al mediodía local, de día, en latitudes más próximas al Sol y en la cresta del ciclo de 11 años de las manchas solares. Por la noche la densidad disminuye hasta un solo estrato F, con densidad pico de electrones menor que el máximo diurno en 1 a 1.5 órdenes de magnitud.

Las frecuencias de radionavegación por satélite son lo suficientemente altas para que no exista refracción o absorción considerable de las trayectorias de los rayos. Sin embargo, debido a las exactitudes que se desean, los pequeños efectos refractivos son importantes y deben tomarse en cuenta en la navegación de precisión.

Los efectos más significativos del paso por la ionosfera en la onda de radio son dos. Primero, la onda se propaga a una velocidad infinitesimalmente menor en la ionosfera que en el vacío. Segundo, cuando se propaga por una región de densidad cambiante la onda se refracta suavemente, dependiendo del gradiente de la densidad. Esto significa que desde un punto situado por encima de la ionosfera hasta uno situado debajo, la onda recorre una trayectoria curva, más larga que la trayectoria recta que conecta los dos puntos. Con ello se producen diferentes efectos sobre un sistema Doppler, tal como el TRANSIT, que sobre un sistema de localización, tal como el GPS.

Debido a que la trayectoria del rayo del satélite al usuario no es recta, la distancia medida por un sistema radiotelemétrico es siempre un poco mayor. Como la trayectoria de los rayos siempre se curva por encima de la trayectoria recta y por tanto "observa" una mayor componente de proyección de la velocidad del satélite, la frecuencia Doppler medida por un sistema Doppler siempre es un poco mayor en magnitud. Sin embargo, éste es un efecto muy pequeño comparado con el que se produce en un sistema Doppler cuyo retardo en la propagación de la trayectoria curva apenas exceda al de la recta. Los efectos en los sistemas TRANSIT y GPS se explican de la manera siguiente.

Dado que la ionosfera crea un retardo mayor que el de una trayectoria recta, la incógnita es cómo calcular la diferencia de retardo.

La flexión de la trayectoria del rayo depende del índice de refracción  $\mu$  o, más bien, de su gradiente a lo largo de la trayectoria. Dicho índice  $\mu$  es función de la frecuencia  $F$  (en hertz) y de la densidad de electrones por centímetro cúbico  $N$ , lo que se expresa como:

$$\mu(N) = \left( 1 - \frac{80.6N}{F^2} \right)^{1/2} \quad (54-18)$$

Cuando  $\mu$  se vuelve imaginario, el gradiente se vuelve infinito y una onda de radio no puede penetrar más la ionosfera. No obstante, para las frecuencias de navegación por satélite y densidades electrónicas pico relativamente grandes,  $\mu$  es muy cercano a 1 y ocurre una flexión mínima de la trayectoria del rayo. Para valores de la frecuencia de 400 y 1 575 MHz, respectivamente, los índices  $\mu$  son del orden de:

$$\mu_{400} = 1 - 8\pi \times 10^{-11} \quad (54-19)$$

$$\mu_{1575} = 1 - \frac{\pi}{2} \times 10^{-11}$$

Tabla 54-1. Cálculos de las órbitas de satélites

Cálculo	Explicación
$\mu = 3.986008 \times 10^{14}$	Parámetro gravitacional del WGS-72 <sup>a</sup> , m <sup>3</sup> /s <sup>2</sup>
$\dot{\Omega}_e = 7.292115147 \times 10^{-5}$	Velocidad de rotación de la Tierra del WGS-72, rad/s
$C_{rc}, C_{rs}, C_{uc}, C_{us}, C_{ic}, C_{is}$	Coefficientes armónicos de zona, dados
$t_{oe}$	Tiempo de efemérides, dado
$e$	Excentricidad, dada
$a$	Semieje mayor, dado
$n_0 = (\mu/A^3)^{1/2}$	
$n = n_0 + \Delta n$	$\Delta n$ dado
$M = M_0 + n(t - t_{oe})$	$M_0$ dado
$M = E - e \sin E$	Implicidad resuelta para $E$
$\cos v = (\cos E - e)/(1 - e \cos E)$	
$\sin v = (1 - e^2)^{1/2} \sin E/(1 - e \cos E)$	
$v = \arctan(\sin v / \cos v)$	
$\phi = v + \omega$	$\omega$ dada
$u = \phi + \delta u$	$\delta u = C_{uc} \cos 2\phi + C_{us} \sin 2\phi$
$r = a(1 - e \cos E) + \delta_r$	$\delta_r = C_{rc} \cos 2\phi + C_{rs} \sin 2\phi$
$x'' = r \cos u$	
$y'' = r \sin u$	
$i = i_0 + \delta_i$	$\delta_i = C_{ic} \cos 2\phi + C_{is} \sin 2\phi$ , $i_0$ dado
$\Omega = (\Omega_0 - \dot{\Omega}_e \cdot t_{oe}) + (\dot{\Omega} - \dot{\Omega}_e) \cdot (t - t_{oe})$	$\Omega_0, \dot{\Omega}$ dado
$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \Lambda(\omega, \Omega, i) \cdot \begin{bmatrix} x'' \\ y'' \\ 0 \end{bmatrix}$	$\Lambda(\cdot)$ dada por la ecuación 54-13

<sup>a</sup> WGS-72, sistema geodésico mundial de 1972.

De la misma manera, la separación angular de las trayectorias recta y curva es muy pequeña. Para una distancia del satélite al usuario de 2 000 km (sistema TRANSIT) y una diferencia de retardo de 10 m, el ángulo de separación de la trayectoria en el satélite es del orden de  $\theta = 4 \times 10^{-3}$  rad, o, lo que es lo mismo, 14 min de arco.

Un modelo de primer orden de la diferencia de distancia es:

$$\Delta R = \frac{K}{F^2} (\text{TEC}) \quad \text{m} \quad (54-20)$$

$$\text{donde } K = \frac{e^2}{4\pi^2 m \epsilon_0} = 80.6 \frac{\text{m}^3}{\text{electrón} \cdot \text{s}^2}$$

$$e = \text{carga del electrón} = 1.6 \times 10^{-19} \text{C/electrón}$$

$$m = \text{masa del electrón} = 9.1 \times 10^{-31} \text{kg/electrón}$$

$$\epsilon_0 = \text{permitividad} = 10^{-9}/36 \pi \frac{\text{C}^2 \cdot \text{s}^2}{\text{m}^3 \cdot \text{kg}}$$

$$F = \text{frecuencia, ciclos/s o Hz}$$

TEC = contenido total de electrones integrados en la columna con 1 m<sup>2</sup> de base a lo largo de la trayectoria desde el usuario hasta el satélite, electrones/m

Es posible obtener el valor de TEC integrando la densidad de electrones por metro cúbico a lo largo de la trayectoria del rayo. Debido a que la diferencia que existe entre las trayectorias curva y recta resulta pequeña, es suficiente integrar a lo largo de la trayectoria recta.

Por lo general, los valores típicos de TEC varían de  $10^{16}$  a  $10^{19}$ .

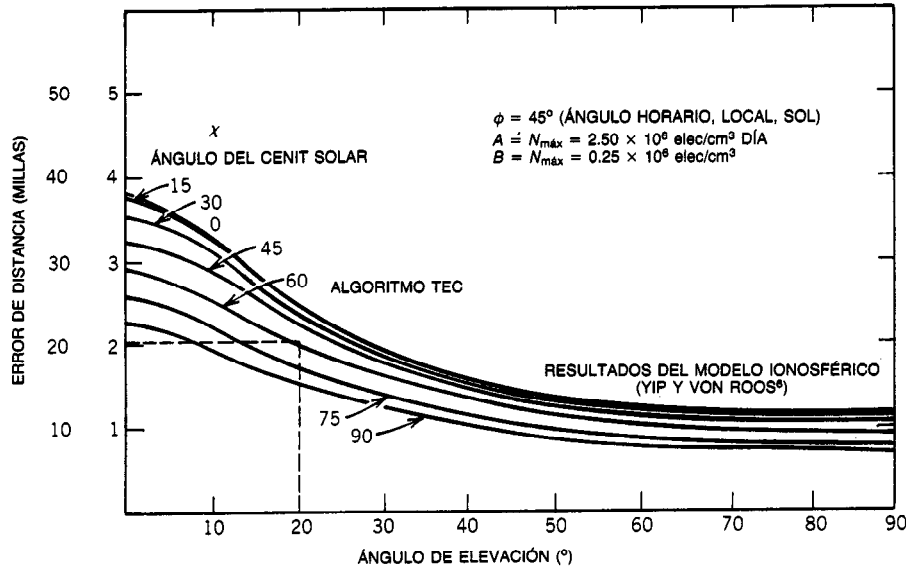
El modelo de densidad de electrones como función de  $r$ , el radio geocéntrico al satélite, ha sido determinado por Yip y Von Roos (véase la referencia 5) con la siguiente expresión:

$$N(h) = N_{\max} \exp(1/2) \left\{ 1 - z + \int_{-\infty}^z \frac{\exp(-x) dx}{[1 - (y_0^2 + z_0^2)/(Hx + R_e - h_{\max})^2]^{1/2}} \right\} \quad (54-21)$$

**Tabla 54-2. Relaciones entre el sistema geocéntrico fijo en la tierra (ECEF) y el sistema de latitud, longitud y altitud (LLA)**

Cálculo	Comentario
$a = 6378135 \text{ m}$	Semieje mayor del WGS-72 <sup>a</sup>
$e^2 = 2f - f^2$	$f = 1/298.26$
$r_c = a/(1 - e^2 \sin^2 \phi)^{1/2}$	Radio de curvatura
$x = (r_c + h) \cos \phi \cos \lambda$	
$y = (r_c + h) \cos \phi \sin \lambda$	
$z = [r_c(1 - e^2) + h] \sin \phi$	

<sup>a</sup> WGS-72, sistema geodésico mundial de 1972.



**Fig. 54-3.** Error de medición ionosférica. TEC = densidad de electrones en la columna.

donde  $N_{\text{máx}}$  = densidad electrónica pico ( $\text{m}^3$ )<sup>-1</sup>, a la altura  $h_{\text{máx}}$

$$z = \frac{r - R_e - h_{\text{máx}}}{H} = \text{altura normalizada}$$

$r$  = radio geocéntrico

$R_e$  = radio de la Tierra

$H$  = altura de escala de la ionosfera, que va desde unos 8 a 100 km de altitud hasta 110 a 700 km de altitud relativa

$\chi$  = ángulo del cenit solar

$y_0$  = coordenadas del usuario en el sistema  $x_0, y_0, z_0$ , donde  $z_0 = [X_0, 0, 0]$  es el subpunto solar

Aplicando un factor de corrección a la ecuación 54-21 para el ángulo de elevación del satélite ( $\gamma$ ) y combinándola con la ecuación 54-20, se obtiene la integral de la diferencia de retardo de la trayectoria:

Esta integral requiere la evaluación numérica por computadora mediante alguna técnica tal como la regla de Simpson, por ejemplo.

Las gráficas de los resultados se trazan sin dificultad, con el ángulo de elevación del satélite  $\gamma$  como abscisa y la diferencia de distancia  $\Delta R$  como ordenada. Se pueden hacer gráficas para valores fijos de ángulo del cenit solar  $\chi$ , o de ángulo horario local  $\phi$ , donde el otro sirve como parámetro de una familia de gráficas. En la figura 54-3 se ilustra un ejemplo con dos valores de  $N_{\text{máx}}$  para un ángulo horario local de +45°, que es aproximadamente las 15:00 horas, hora local. Se ilustra también una familia de curvas para el ángulo del cenit solar  $\chi$  desde 0 hasta 90°.<sup>6</sup>

Una característica interesante del ejemplo que se da es que la diferencia de retardo no aumenta en forma considerable para ángulos de elevación menores, como se había afirmado. Se observa que el incremento del

$$\Delta R = \left( \frac{e^2}{4\pi^2 m \epsilon_0 F^2} \right) \int_S N_{\text{máx}} \cdot \exp \left\{ \frac{1 - z + \int_{-\infty}^z \frac{\exp(-x) dx}{[1 - (y_0^2 + z_0^2)/(Hx + R_e + h_{\text{máx}})^2]^{1/2}}}{2[1 - (R_e^2 \cos^2 \gamma)/r^2]^{1/2}} \right\} dr \quad (54-22)$$

ángulo de elevación desde 0 hasta 90° sólo ocurre en un factor de 2 a 3. Se aprecia que el valor de  $\Delta R$  es linealmente dependiente de  $N_{\text{máx}}$ .

Dada la existencia de la diferencia de retardo en la propagación, se suscitan dos preguntas:

1. ¿Cuál es el efecto sobre la solución de navegación?
2. ¿Qué se puede hacer al respecto?

A fin de contestarlas correctamente, en primer lugar se identificará el efecto.

Tanto en el sistema TRANSIT como en el GPS, el efecto de  $\Delta R$  consiste en un sesgo que aleja del satélite la posición medida. Si se pueden recibir rápidamente señales de tres satélites con el mismo ángulo de elevación aproximadamente, pero separados en acimut 120°, entonces los sesgos de  $\Delta R$  tenderán a cancelarse en las coordenadas de latitud y longitud. Sin embargo, permanecerá el efecto sobre la altitud, y habrá la tendencia a medir una altitud menor que la real. En lo que respecta a la navegación de superficie este hecho no constituye un problema.

Con respecto a la corrección del efecto, existen varias posibilidades. La primera consiste en emplear dos frecuencias de transmisión individuales para corregir los efectos de la diferencia de retardo. En esta técnica se aprovecha el que esta diferencia es inversamente proporcional al cuadrado de la frecuencia.

Supóngase que se utilizan dos frecuencias  $f_1$  y  $f_2$  para la transmisión, donde  $f_1 < f_2$ , y las frecuencias se generan coherentemente en fase a partir de la misma fuente osciladora, de manera que:

$$\frac{f_2}{f_1} = (K)^{1/2} \quad (54-23)$$

donde  $K$  se conoce y es estable. Ahora, las distancias medidas a las frecuencias  $f_1$  y  $f_2$  se representan respectivamente como:

$$\begin{aligned} R_1 &= R + \Delta R_1 \\ R_2 &= R + \Delta R_2 \end{aligned} \quad (54-24)$$

donde  $R$  es la distancia real y  $\Delta R_1$ ,  $\Delta R_2$  son las diferencias de distancia a las frecuencias  $f_1$  y  $f_2$ . Existe una constante  $C_1$  tal que:

$$\Delta R_1 = \frac{C_1}{f_1^2}, \quad \Delta R_2 = \frac{C_1}{f_2^2} \quad (54-25)$$

por tanto:

$$\Delta R_1 = \left(\frac{f_2}{f_1}\right)^2 \cdot \Delta R_2 \quad (54-26)$$

y:

$$R_1 - R_2 = \Delta R_1 - \Delta R_2 \quad (54-27)$$

A partir de estas ecuaciones se deduce que:

$$R = \frac{R_2(f_2/f_1)^2 - R_1}{(f_2/f_1)^2 - 1} = \frac{KR_2 - R_1}{K - 1} \quad (54-28)$$

donde  $K$  está dada por la ecuación 54-23.

Por tanto, a partir de la ecuación 54-28 puede determinarse la distancia real  $R$  midiendo las distancias aparentes  $R_1$  y  $R_2$  a las dos frecuencias conocidas,  $f_1$  y  $f_2$ , cuya razón es una constante estable,  $\sqrt{K}$ .

En el caso de la transmisión Doppler, sea  $D$  el desplazamiento Doppler de la frecuencia  $f$ , donde:

$$D = \frac{v_R}{C} \cdot f \quad (54-29)$$

y  $v_R$  es la componente radial de la velocidad (con signo) y  $C$  la velocidad de la luz. Derivando con respecto al tiempo la ecuación 54-28 se obtiene:

$$\frac{\dot{R}}{C} f_1 = \frac{D_2(f_2/f_1) - D_1}{(f_2/f_1)^2 - 1} = \frac{\sqrt{K} D_2 - D_1}{K - 1} \quad (54-30)$$

La cantidad  $(\dot{R}/C)f_1$  es el desplazamiento Doppler real a la frecuencia  $f_1$ .  $D_1$  y  $D_2$  son los desplazamientos Doppler aparentes medidos a las frecuencias  $f_2$  y  $f_1$ , respectivamente.

Las ecuaciones 54-28 y 54-30 demuestran el método de las dos frecuencias para corregir mediciones de distancia (GPS) o de desplazamiento Doppler (TRANSIT) a fin de considerar los efectos de la diferencia de distancia inducida por la ionosfera.

Existe otra posibilidad para corregir  $\Delta R$  sin utilizar frecuencias múltiples. Como se mostrará más adelante con respecto a TRANSIT y GPS, debe haber tantas mediciones independientes de datos del satélite como variables de navegación por despejar. Por tanto, para determinar la posición del usuario  $x$ ,  $y$ ,  $z$  se requieren tres mediciones independientes. En el caso de que el reloj del usuario tenga un sesgo respecto al tiempo del satélite, se necesita una cuarta medición para resolver ese sesgo. Dadas unas estimaciones de la posición del usuario, la posición del satélite y la posición del Sol (todas calculables), la única incógnita para encontrar  $\Delta R$  con la ecuación 54-22 es  $N_{\text{máx}}$ , la densidad electrónica pico. Si se realizaran cinco mediciones independientes del satélite con trayectorias de rayos que pasaran todas por una región ionosférica de densidad común (el mismo valor de  $N_{\text{máx}}$ ), entonces en teoría también podría despejarse  $N_{\text{máx}}$ . Esta solución requeriría muchos cálculos, pero no es imposible. Este método tiene un potencial a futuro.

#### 54.6 TRANSIT, SISTEMA DOPPLER DE BAJA DINÁMICA

El sistema de satélite TRANSIT emplea múltiples satélites en órbitas polares circulares con período de 107 minutos. La altitud orbital nominal es de 1 075 km. El sistema fue implementado originalmente por la Armada de Estados Unidos para uso exclusivo del Department of Defense. Entró en operación en 1964 y fue

entregado para su uso civil en 1967. Los satélites en órbita son repuestos en caso necesario por otros del tipo original que se tienen almacenados. Asimismo, un programa de mejora del sistema TRANSIT ha producido un satélite mejorado compatible con el TRANSIT, llamado NOVA, que ahora se tiene a disposición. Los nuevos satélites emplean el control orbital para mantener la precesión del plano orbital (ascensión recta del nodo ascendente) en valores insignificantes. La precesión orbital de los cinco satélites originales de TRANSIT era un problema que en algunos casos producía tiempos de espera inaceptablemente largos entre marcaciones o puntos de posición (ref. 3, pág. 48).

Aunque originalmente se diseñó para cuatro satélites con planos separados por  $45^\circ$  de longitud (con los nodos ascendentes separados  $90^\circ$ ), la precesión orbital sin control pronto hizo muy deseable un quinto satélite. Recientemente, el mismo efecto ha llevado a planificar la construcción de un sexto satélite. La imposibilidad de controlar la precesión orbital ha hecho que los tiempos de espera, incluso con seis satélites, pasen de las cuatro horas, con el 10% de probabilidad. Para seis satélites el tiempo medio entre marcaciones o puntos de posición es de 81 minutos en el peor de los casos. Ambos casos se dan en el ecuador.

Se dispone de un punto de posición de TRANSIT cada vez que está visible un solo satélite, lo que ocurre en cualquier sitio dado durante 80 a 100 minutos como promedio en el peor de los casos de latitudes bajas, dependiendo de si están activos cinco o seis satélites. Por tanto, el TRANSIT es un sistema de fijación de la posición (marcación) y no un sistema de navegación continua. Para los usuarios marítimos con una navegación a la estima adecuada y baja dinámica de maniobra, el TRANSIT es un sistema muy útil.

Cada punto de posición se obtiene de un satélite cuando éste pasa por la ubicación del usuario. Un paso de satélite útil será visible cada 10 a 18 minutos. Durante este periodo, el receptor del usuario deberá realizar mediciones Doppler desde el satélite en varias ocasiones. Cada periodo de medición Doppler es en esencia una de las varias mediciones independientes que se deben hacer para resolver las diversas incertidumbres de navegación. Dado que cada medición Doppler de satélite puede corregir la posición estimada del usuario sólo en la dirección de trayectoria óptica del satélite, se puede apreciar que son necesarias varias mediciones a medida que el satélite cambia de posición en acimut con respecto al usuario. Como la altitud del usuario no es importante en el mundo marítimo, bastan dos mediciones adecuadas en acimuts ortogonales para resolver la longitud y la latitud. También se requiere una tercera medición para resolver la incertidumbre en la frecuencia del oscilador de referencia Doppler del usuario, que más tarde entra en el procesamiento de las mediciones. En la práctica, se toman tantas mediciones independientes como sea posible en los tiempos disponibles del satélite a fin de alisar la posición calculada. El equipo moderno realiza de 20 a 40 series de mediciones.

Debido a que las diversas mediciones Doppler se realizan en tiempos distintos, se debe predecir el mo-

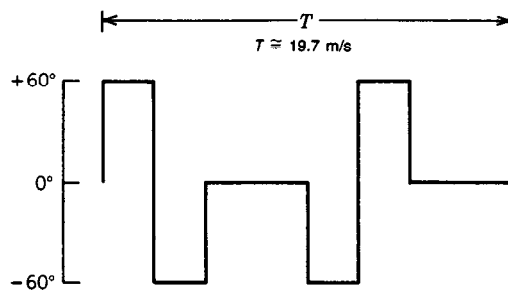


Fig. 54-4. Onda de forma de señal de TRANSIT.

vimiento del usuario a fin de tener una posición estimada en el tiempo de la medición Doppler, que puede corregirse por medio de esta misma medición. Por tanto, se requiere navegación a la estima durante el intervalo de medición. Sobra decir que este requisito limita al usuario en la maniobra dinámica durante este periodo.

#### 54.6.1 Señales de los TRANSIT y estructura del sistema

Los satélites de TRANSIT transmiten dos señales, una a 399.968 MHz y otra a 149.988 MHz, cada una con estabilidad a largo plazo de una parte en  $10^{11}$  durante un tiempo de paso del satélite. Cada frecuencia se desvía 80 partes por millón por debajo de los 400.00 y 150.00 MHz, respectivamente. Esto garantiza que el máximo desplazamiento Doppler positivo no dará por resultado frecuencias recibidas mayores de 400.00 y 150.00 MHz, respectivamente, que son las frecuencias nominales de medición de referencia Doppler del receptor del usuario.

El objetivo de transmitir dos frecuencias es corregir el efecto ionosférico, como se describe con detalle en la sección 54.5. Un receptor TRANSIT de precisión emplea canales duales para la recepción de estas dos frecuencias. Un receptor estándar de un solo canal utiliza únicamente la señal de 400 MHz.

Las potencias de las señales transmitidas son del orden de un watt, dejando un margen significativo para la mayor parte de las condiciones de los canales si se emplean receptores de bajo ruido con técnica de vanguardia. Las señales portadoras se manipulan por defasamiento en cuadratura. Se coloca una onda de forma de tres niveles en la fase seno para transportar los datos binarios del mensaje del satélite. Esta onda de forma, que se ilustra en la figura 54-4, representa un "uno" binario según se muestra. Para un "cero" binario, la onda de forma se invierte. La desviación de fase de la onda de forma de datos es  $\pm 60^\circ$ . Se coloca una onda de forma de reloj de dos niveles en la fase coseno de la señal. El reloj funciona a dos veces la velocidad de transferencia de bits y se utiliza para sincronizar el detector de bits del receptor. La desviación de fase de la portadora es tal que se mantiene una componente portadora residual no modulada, que contiene el 56.25 % de la potencia de señales total. Esto hace posible que el receptor emplee un ciclo sincronizado (enclavado en



fase) estándar para el seguimiento de la componente portadora no modulada. El oscilador de ciclo produce sinusoides de referencia de desmodulación para recuperar los bits de datos y el reloj de fases seno y coseno, respectivamente, así como la señal Doppler.

Las potencias asignadas a los datos y el reloj son el 37.5% y el 6.25%, respectivamente (ref. 3, pág. 48). En la figura 54-5 se presenta una implementación representativa del receptor.

El ciclo de navegación de TRANSIT está organizado en torno a un intervalo básico de dos minutos, que es la duración de los mensajes de datos de navegación transmitidos. El mensaje consta de 6 103 dígitos binarios de datos organizados en 26 cuadros, seguidos de 19 bits finales. Cada cuadro consta de seis palabras de 39 bits. Los 25 bits finales de cada bloque de mensajes de dos minutos son una palabra de sincronización que denota la marca de tiempo entre mensajes de dos minutos. Cada mensaje comienza y termina precisamente en el inicio de un minuto par en el tiempo del sistema TRANSIT (tiempo del satélite).

Los parámetros orbitales están situados en la sexta palabra de los cuadros 1 a 22, inclusive. Esto es, los elementos de la órbita se encuentran en la sexta palabra de cada cuadro del 1 al 22. Las últimas palabras de los cuadros 9 al 22 son los elementos que definen una órbita elíptica suave (lisa) en el periodo aproximado de 12 horas entre cargas ascendentes de la memoria de datos del satélite. Las últimas palabras de los cuadros 1 al 8 inclusive son las correcciones de tiempo real a la órbita suave, manipuladas a intervalos específicos de dos minutos. Es decir, se dan correcciones para los seis minutos anteriores, los dos minutos presentes y los ocho minutos posteriores. Así, mediante la interpolación las correcciones se pueden estimar en un periodo de 16 minutos. Cada dos minutos las entradas en la palabra seis de los cuadros 1 al 9 se hacen retroceder en el tiempo durante dos minutos y se introduce una nueva palabra en el cuadro 9 (véase la referencia 3).

#### 54.6.2 Técnica de localización Doppler de TRANSIT

La componente portadora no modulada es radiada desde el satélite a una frecuencia  $f_i = 399.968$  MHz. Ésta es recibida por el usuario como una frecuencia  $f_r$ , que es  $f_i$  más el desplazamiento Doppler. Como la velocidad orbital de un satélite con periodo de 107 minutos es 6 235 km/h, el máximo desplazamiento Doppler cuando el satélite cruza el horizonte de los  $0^\circ$  es  $\pm 7\,057$  Hz en una portadora de 400 MHz.

En vez de medir directamente la frecuencia Doppler, los cruces por cero Doppler se cuentan como de 23 segundos, bajo el control de las marcas de sincronización inherentes en los datos de telemetría. Debido a la desviación de la frecuencia de  $8 \times 10^{-6} f_i$  en el satélite ( $\sim 32$  kHz), la frecuencia  $f_i$  que se cuenta siempre refleja un sesgo Doppler positivo, aun cuando el satélite se aleje del usuario. Esta desviación más cualquier inexactitud a largo plazo en la frecuencia del oscilador del usuario debe resolverse junto con las coordenadas de posición del usuario.

El oscilador del receptor del usuario genera una frecuencia de referencia  $fg = 400.00$  MHz. La salida de frecuencia Doppler del receptor es entonces  $fd = fg - fr$ . Esta frecuencia Doppler se cuenta desde el momento  $t_1$  hasta el  $t_2$ , un periodo de exactamente dos minutos en el satélite. La cuenta resultante  $m$  está dada como:

$$m = \int_{t_1}^{t_2} fd \, dt = \int_{t_1}^{t_2} (fg - fr) \, dt \quad (54-31)$$

Ahora, los momentos  $t_1'$  y  $t_2'$  son tiempos de eventos de señales medidos por el reloj del usuario. Se pueden relacionar con los tiempos de eventos de señales del satélite  $t$  por medio de:

$$t' = t + \frac{R_{t,t'}}{C} \quad (54-32)$$

Es decir, el evento de señales del satélite observado en el momento  $t$  (como una marca de sincronización de datos) se aprecia en el receptor en el momento  $t'$ .  $R_{t,t'}$  representa la distancia que recorrió la señal entre  $t$  y  $t'$ .

Por tanto, la cuenta Doppler puede expresarse en la forma:

$$m = \int_{t_1 + R_1/C}^{t_2 + R_2/C} (fg - fr) \, dt = \int_{t_1 + R_1/C}^{t_2 + R_2/C} fg \, dt - \int_{t_1 + R_1/C}^{t_2 + R_2/C} fr \, dt \quad (54-33)$$

La primera integral es simple, ya que  $fg$  es constante durante el periodo de integración. Sin embargo  $fr$  varía durante este periodo. Con todo, se cumple un atrayente argumento de Stansell<sup>3</sup>, según el cual el número de ciclos contados en el periodo entre las dos marcas de sincronización recibidas debe ser idénticamente igual al número transmitido entre las marcas. De esta forma se tiene:

$$\int_{t_1 + R_1/C}^{t_2 + R_2/C} fr \, dt = \int_{t_1}^{t_2} ft \, dt = f_i(t_2 - t_1) \quad (54-34)$$

Con la ecuación 54-34  $m$  se convierte en:

$$m = (fg - f_i) \cdot (t_2 - t_1) + \frac{fg}{C} (R_2 - R_1) \quad (54-35)$$

En la ecuación 54-35 se observa que la medición de navegación  $m$  consta de una distancia diferencial  $(R_2 - R_1)$ , ponderada por el factor  $fg/C$ , más la cantidad  $(fg - f_i) \cdot (t_2 - t_1)$ .

Las mediciones de navegación de TRANSIT, análogas a la forma de la ecuación 54-2, son en esencia mediciones de distancia diferencial. Siguiendo la trayectoria de las ecuaciones 54-2 a 54-9, se busca una función  $f(x, y, z)$  relacionada con la medición  $m$ . Diversos fabricantes de receptores TRANSIT han seguido este procedimiento de distintas formas, pero el siguiente análisis es lo suficientemente representativo para fines explicativos.

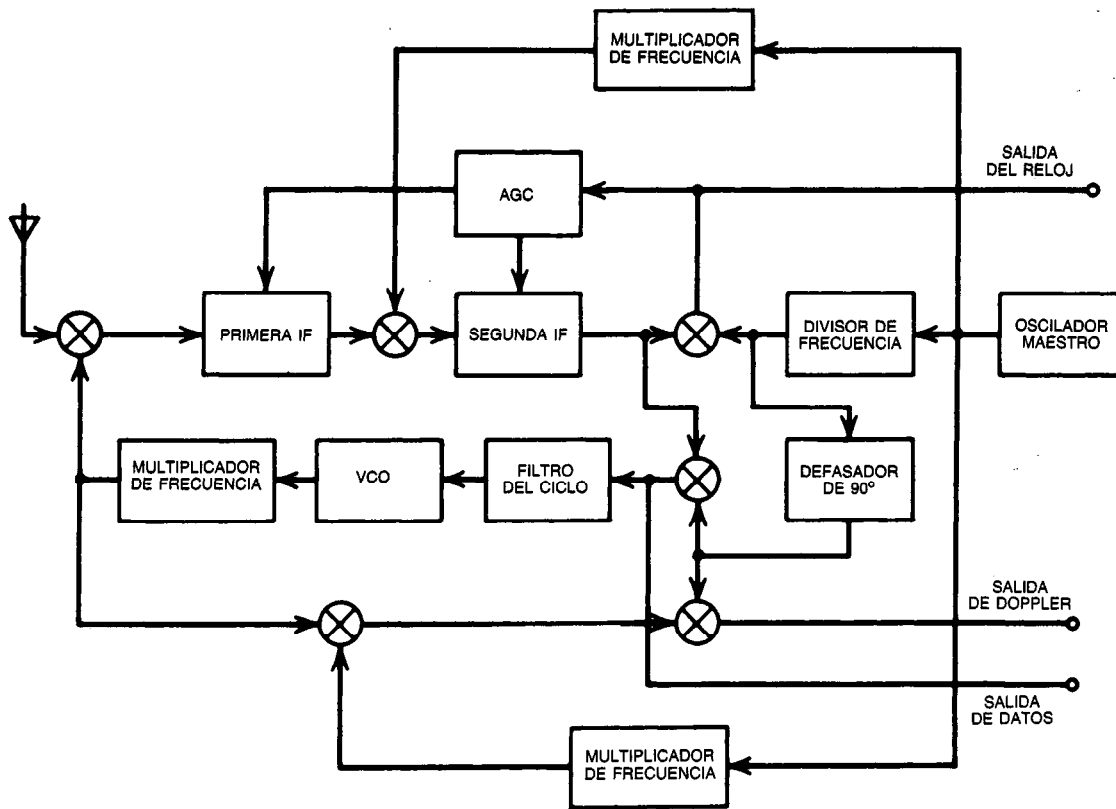


Fig. 54-5. Implementación del receptor de TRANSIT. AGC = control de ganancia automático; VCO = controlado por voltaje.

La posición real del usuario en el sistema cartesiano ECEF se representa como  $[x, y, z]^T$ . La relación de distancia se expresa en términos generales como:

$$R_i[(x_{si} - x)^2 + (y_{si} - y)^2 + (z_{si} - z)^2]^{1/2} = f_i(x, y, z) \quad (54-36)$$

Donde  $i$  denota la  $i$ -ésima medición y  $(x_{si}, y_{si}, z_{si})$  representa la posición calculable del satélite para esa medición. Así:

$$R_2 - R_1 = f_2(x, y, z) - f_1(x, y, z) \quad (54-37)$$

Entonces se tiene:

$$m = (fg - ft)(t_2 - t_1) + \frac{fg}{C}[f_2(x, y, z) - f_1(x, y, z)] \quad (54-38)$$

Si la cantidad  $(fg - ft)(t_2 - t_1)$  se considera un sesgo de la diferencia de distancia desconocida  $b$ , también puede optarse por resolverlo, además de las coordenadas ECEF  $x, y, z$ . Obsérvese que  $fg$ , que se deduce de la referencia de frecuencia básica del usuario, se puede suponer conocida. Pero  $(fg - ft)$  no se conoce *a priori*, ya que  $ft$  no está disponible para medirse con respecto a  $fg$ . Por tanto, es posible incorporar  $b$  en el estado de navegación y reexpresar la ecuación 54-38 como:

$$m = f(x, y, z, b) \quad (54-39)$$

Ahora, el siguiente paso usual es la linealización del problema suponiendo una posición  $(x_0, y_0, z_0, b_0)$ , quizá por navegación a la estima. Esta posición se toma en el punto medio del curso del usuario que presumiblemente se recorre durante el periodo de medición completo del satélite<sup>7</sup>.

La linealización se expresa como:

$$\begin{aligned} e &= m - f(x_0, y_0, z_0, b_0) = m - m_0 = \\ &= \left. \frac{\partial f}{\partial x} \right|_{x_0} \cdot (x - x_0) + \left. \frac{\partial f}{\partial y} \right|_{y_0} \cdot (y - y_0) + \left. \frac{\partial f}{\partial z} \right|_{z_0} \cdot (z - z_0) + \left. \frac{\partial f}{\partial b} \right|_{b_0} \cdot (b - b_0) \end{aligned} \quad (54-40)$$

Se pueden realizar cuatro mediciones de este tipo, lo que da por resultado:

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \Big|_{x_0} & \frac{\partial f_1}{\partial y} \Big|_{y_0} & \frac{\partial f_1}{\partial z} \Big|_{z_0} & \frac{\partial f_1}{\partial b} \Big|_{b_0} \\ \frac{\partial f_2}{\partial x} \Big|_{x_0} & . & . & . \\ \frac{\partial f_3}{\partial x} \Big|_{x_0} & . & . & . \\ \frac{\partial f_4}{\partial x} \Big|_{x_0} & . & . & \frac{\partial f_4}{\partial b} \Big|_{b_0} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \\ b - b_0 \end{bmatrix} \quad (54-41)$$

En la ecuación 54-41, los términos  $f_i$  son las versiones de la ecuación 54-38 que se calculan para la posición del satélite aplicable en el  $i$ -ésimo tiempo de medición. La ecuación 54-41 puede expresarse en forma compacta como:

$$(\mathbf{m} - \mathbf{m}_0) = H(\mathbf{x} - \mathbf{x}_0) \quad (54-42)$$

donde  $\mathbf{m}$  es el vector de cuatro mediciones,  $\mathbf{x}$  es el vector de cuatro coordenadas y  $H$  es una matriz de  $4 \times 4$  de derivadas parciales de la transformación, conocida en general como matriz jacobiana.

Para un paso de satélite TRANSIT, la matriz  $H$  puede invertirse, con la inversa  $H^{-1}$ . Por tanto la ecuación 54-42 puede reexpresarse como:

$$\mathbf{x} - \mathbf{x}_0 = H^{-1}(\mathbf{m} - \mathbf{m}_0) \quad (54-43)$$

Ahora, cada solución requiere sólo cuatro mediciones de cuenta Doppler de 23 segundos. Es posible realizar de 20 a 40 mediciones durante un paso nominal. Por tanto, existe la capacidad de suavizar o filtrar las posiciones. Si el usuario está en reposo, se logra promediando las coordenadas. Si está en movimiento, es posible utilizar un filtro secuencial del tipo Wiener o Kalman. Incluso es posible operar un filtro secuencial con mediciones individuales, ya que cada residuo,  $e_i$ , puede "proyectarse" en las cuatro coordenadas mediante los elementos de la matriz  $H^{-1}$ .

Esta explicación teórica se dio sólo con fines ilustrativos. Los cálculos que se realizan en el receptor de un fabricante pueden variar en detalle dentro del ámbito de la técnica de la telemetría diferencial básica. Asimismo, algunos fabricantes calculan la latitud, longitud y sesgo (error) directamente a partir de los residuos, transformando las coordenadas  $x, y, z$  sólo para calcular las distancias diferenciales estimadas. Se pueden hacer diversos refinamientos para incrementar la exactitud de un usuario en movimiento.

### 54.6.3 Rendimiento y características de error del TRANSIT

La exactitud de un punto de localización navegacional típico del TRANSIT depende del tipo de receptor que

se utilice, del movimiento del usuario y la configuración geométrica del paso del satélite dado. Stansell<sup>3</sup> informa exactitudes típicas para un usuario estacionario de 27 a 37 m rms en el caso de un equipo de doble canal (corregido para considerar el efecto de la ionosfera) y de 80 a 100 m rms en el caso de un equipo de canal único. La regla empírica sugerida para usuarios en movimiento es 370 m de error de posición por cada nudo de velocidad desconocida del usuario.

Para un usuario estacionario con doble canal, la mayor parte del error es sistemático y no inducido por instrumentación. Stansell<sup>3</sup> cita un presupuesto de error de 1973, atribuido al Applied Physics Laboratory, que asignó de 15 a 30 m rms a las incertidumbres en el cálculo de posiciones de satélites y sólo de 3 a 6 m rms al ruido de la medición. Para un usuario diurno con un solo canal, los errores inducidos ionosféricamente a 400 MHz fueron algunas veces tan grandes como de 200 a 500 m, principalmente en longitud. Los errores debidos a la refracción troposférica a ángulos de elevación mayores de unos 20° típicamente se cotizaron en 4 a 8 m. En el caso de equipos que no calculan la altitud como parte de la solución de navegación, los errores de altitud desconocidos se propagan en la solución de navegación principalmente como errores en la longitud. En el caso de estos equipos se debe conocer la altitud con respecto al geoide WGS-72, en vez de al nivel del mar. Es bien sabido que la superficie oceánica física presenta "crestas" y "valles" de  $\pm 80$  a 100 m.

Debido a su sensibilidad a movimientos desconocidos del usuario y su tiempo medio de 81 minutos entre puntos de localización, el sistema TRANSIT se ha caracterizado como un sistema de fijación de la posición para usuarios de baja dinámica. Ésta no es una clasificación despectiva, ya que el TRANSIT hizo avanzar la ciencia de la navegación y la geodesia en un orden de magnitud después de que llegó a ser operacional en 1964. Sin embargo, el hardware de computación miniaturizado se ha desarrollado lo suficiente desde entonces, para hacer posible otro avance en orden de magnitud, el NAVSTAR.

## 54.7 GPS NAVSTAR, SISTEMA TELEMÉTRICO DE ALTA DINÁMICA

El sistema de localización global (GPS) NAVSTAR es un sistema en el que se ha llegado al empleo de 18 satélites en órbitas de 12 horas con 55° de inclinación. El sistema está siendo implementado por el Department of Defense de Estados Unidos para uso militar. Sin embargo, tiene un canal de "libre acceso" (C/A) disponible para el uso civil general. El programa de desarrollo del GPS surgió de dos programas de la década de 1960, el programa 621-B de la fuerza aérea y el programa Timation de la marina, que se fusionaron en 1973. El GPS es un sistema de navegación por satélite de la segunda generación en el que se aplica la tecnología telemétrica deseudoruido (PN), creada por la NASA<sup>8</sup>, a la tecnología de navegación por satélite integrada en el sistema TRANSIT.

En el GPS se emplean satélites que están controlados con precisión en sus posiciones orbitales. En realidad, basta el conocimiento de un conjunto de elementos orbitales o efemérides de un año para predecir tiempos de visibilidad de satélites con un margen de cinco minutos a partir de una posición conocida en la Tierra.

Al igual que el sistema TRANSIT, el GPS es un sistema de navegación pasivo en lo que corresponde al usuario, ya que sólo la recepción de señales transmitidas por el satélite es utilizada por el navegante para calcular la posición. A diferencia de lo que ocurre en el caso del sistema TRANSIT, en el GPS se emplea la recepción simultánea o casi simultánea de señales de cuatro satélites para calcular tres coordenadas de posición y una diferencia de tiempo, debida al error entre los relojes del satélite y del usuario. Por tanto, dada la visibilidad de cuatro satélites, GPS ofrece la posibilidad del procesamiento de navegación continua en vez de puntos de posición aislados interconectados con la navegación a la estima. Como se describirá, los tiempos de integración de mediciones son mucho menores con el GPS que con el TRANSIT. De este modo se pueden determinar posiciones en el orden de una cada segundo. Por consiguiente, el GPS ofrece la posibilidad de "seguir" vehículos caracterizados por maniobras de alta dinámica.

Las posiciones de los satélites GPS se distribuyen uniformemente en torno a la Tierra, tres por órbita, en seis órbitas\*. Con la constelación orbital completa de 18 satélites, hay más de cuatro visibles normalmente en un momento dado cualquiera en cualquier parte de la Tierra. Esta circunstancia conlleva el problema, que se analiza más adelante, de elegir los cuatro satélites que tengan la configuración geométrica de navegación más "sólida".

Asimismo, si un vehículo usuario llega a perder la línea de visión con un satélite del GPS debido, por ejemplo, al bloqueo de la señal por la superestructura del vehículo, existe la posibilidad de "adquirir" de inmediato otro satélite para sustituir al que se acaba de perder. Este procedimiento es el mismo que cuando uno de cuatro satélites se coloca por debajo del horizonte utilizable. Esto último se puede predecir, desde luego, y es posible planificar por anticipado la adquisición de un satélite de repuesto.

El procesador de navegación de GPS opera a partir de mediciones de distancia entre el satélite y el usuario. En realidad la medición es la del tiempo que transcurre entre el momento en que el satélite transmite una referencia de señal conocida y el momento en que el usuario recibe esa misma referencia. Dada la velocidad de la luz, la medición del tiempo transcurrido es igual a una distancia medida. Entonces se mide el tiempo de transmisión del satélite según su reloj, que se pone exactamente a la hora del GPS, con un error del orden de 3 ns. No obstante, el tiempo de recepción del usuario se mide con respecto a su propio reloj con base en un oscilador de frecuencia de referencia del usuario. El

oscilador del usuario suele ser de mucho menor calidad que los patrones de frecuencia para satélites en muchos órdenes de magnitud.

Los satélites de demostración del GPS, la llamada serie NDS, han transportado patrones de frecuencia de cesio y rubidio con estabilidades de frecuencia básicas a largo plazo del orden de  $10^{-12}$  o  $10^{-13}$  ( $\Delta f/f_0$ ). Ésta es una estabilidad verdaderamente sorprendente si se considera que la desviación de frecuencia debida sólo a efectos de relatividad es del orden de cuatro partes en  $10^{-10}$ .<sup>9</sup> Por tanto, los satélites pueden mantener un tiempo de referencia del sistema en un intervalo de varios nanosegundos en un periodo de 12 horas, que es casi igual a 1 m de distancia. Los osciladores de referencia de frecuencia del usuario (suponiendo un equipo civil de costo razonable) tienen estabilidades de frecuencia a largo plazo del orden de  $10^{-7}$  a  $10^{-9}$ . Esta estabilidad produce derivas del reloj de 1 a 100 ns/s, que equivalen a 0.3 a 30 m de distancia por segundo. Por tanto, es necesaria la inclusión de una cuarta coordenada en la posición del usuario, a saber, el sesgo (desviación) de su reloj (en metros).

Debido a que la distancia medida siempre contiene un error debido a la deriva del reloj, la cantidad medida se denomina pseudodistancia. Las pseudodistancias sin corrección se procesan directamente, como se describirá más adelante, para despejar las cuatro coordenadas del usuario.

#### 54.7.1 Señales y estructura del sistema GPS

El GPS opera con dos frecuencias de señales disponibles del satélite, como lo hace el sistema TRANSIT, a fin de eliminar el tiempo de retardo extra desconocido por efecto ionosférico. Estas frecuencias son 1575.420 MHz, llamada L1, y 1227.600 MHz, llamada L2. Estas dos frecuencias de ondas portadoras son generadas coherentemente en fase por la multiplicación de frecuencia de la misma frecuencia estándar básica de 10.23 MHz. Los factores de multiplicación son 154 para L1 y 120 para L2. En realidad, la referencia básica del satélite se reduce de 10.23 MHz en  $4.45 \times 10^{-10} \times 10.23$  MHz para ecualizar el efecto de la relatividad, pero no es importante para nuestra consideración de la estructura de las señales.

La frecuencia portadora transmitida L1 transporta dos señales en cuadratura de fase. Una fase transporta la señal telemétrica pseudoruido (PN) de libre acceso (C/A) y un enlace de datos de 50 bauds. La fase ortogonal transporta una señal telemétrica de PN de precisión (el código P) y el mismo enlace de datos de 50 bauds. La relación de fase exacta es:

$$\text{Fase portadora de C/A} = \text{Fase portadora P} + 90^\circ \quad (54-44)$$

La portadora transmitida L2 puede transportar la señal de C/A o el código P.

Las señales telemétricas y la señal del enlace de datos son digitales, de diversas velocidades de transmisión (en bauds). En el caso del código P dicha velocidad es 10.23 megabauds, en el del código C/A es 1.023

\* Según los planes revelados por la U.S. Air Force Space Division en el momento en que esto se escribe.

megabauds, y para el enlace de datos de 50 bauds. Los bits del enlace de datos se combinan con los bits del código por adición módulo dos (OR excluyente). El flujo compuesto de bits de datos de código se modula en la fase portadora adecuada mediante manipulación por defasamiento binaria completa ( $\pm 90^\circ$ ), que no deja residuo de portadora no modulada en el espectro transmitido. Esto es, no se transmite una referencia de fase de la portadora para las diversas señales digitales. Por tanto, es de esperar que algunos de los problemas principales en la recepción de estas señales tengan que ver con el logro de la sincronización de las señales.

Los códigos de telemetría PN son ejemplos de secuencias de registros de desplazamiento lineal de longitud máxima<sup>10</sup>, llamados códigos Gold en honor de su inventor, Robert Gold. Las estructuras de los códigos PGS Gold específicos se explican en detalle en la referencia 11. La longitud no ambigua (o periodo de repetición) de los códigos C/A es 1 023 chips (bits), o 1.0 ms, exactamente. Este periodo de tiempo básico se relaciona asimismo con el periodo del reloj de los bits de datos, ya que hay exactamente 20 épocas de código C/A por cada bit de datos. Como el código C/A tiene solamente 1.0 ms de largo, puede resolver distancias en forma inequívoca sólo en múltiplos de alrededor de 300 km. La resolución de la ambigüedad se realiza durante el proceso inicial de adquisición de satélites como parte de la obtención del primer punto de posición.

En realidad sólo existe un código P, que es el producto de dos códigos PN cuyas longitudes en chips son primos relativos (es decir, sin divisores comunes). Estos dos códigos tienen longitudes de 15 345 000 y 15 345 037 chips, respectivamente. Como un periodo de chip es 100 ns, la longitud del código del producto es mayor de 38 semanas. Los códigos P de satélites individuales se toman como segmentos de una semana (sin superposición) del código de 38 semanas de longitud. El código largo se vuelve a iniciar a medianoche entre el sábado y el domingo (tiempo de Greenwich), cada semana.

Los datos digitales que transmite cada satélite contienen todos los parámetros orbitales keplerianos para los cálculos de la posición del satélite que se muestran en la tabla 54-1. Además, en un mensaje de datos del satélite hay elementos orbitales de exactitud reducida para calcular en forma aproximada las posiciones de los otros satélites. La información orbital precisa se denomina efemérides, mientras que los datos menos precisos se conocen como almanaque.

En la transmisión de datos por satélite se emplean palabras de 30 bits, con 10 palabras por cada subcuadro de seis segundos. Cinco subcuadros conforman un mensaje de datos del satélite. Con 20 ms/b, cada palabra dura 0.6 segundo, cada subcuadro emplea seis segundos y un mensaje completo requiere 30 segundos. Todos y cada uno de los mensajes de datos de 30 segundos contienen el almanaque para cada uno de los 18 satélites posibles. Por tanto, para adquirir el almanaque de todos los satélites se necesitan nueve minutos.

Cada subcuadro de seis segundos comienza con una palabra de telemetría de 30 bits (TLM). Los primeros

ocho bits de la palabra TLM son el "8B" hexadecimal. Debido a la forma en que se detectan los datos en un receptor, los bits se pueden complementar. Los dos bits finales de la palabra TLM deben ser ceros. Si son unos, se invierten los datos. Por tanto, una búsqueda de "8B" o su complemento establece la sincronización de los subcuadros.

La segunda palabra de 30 bits de cada subcuadro es una palabra de entrega (HOW). Esta palabra es un número (cuenta  $z$ ), el cual, cuando se multiplica por 4, indica el número de épocas de 1.5 segundos (épocas X1) que habrán ocurrido desde el inicio de la semana GPS, al inicio del siguiente subcuadro (palabra TLM). El objetivo de esta palabra es hacer posible la sincronización aproximada del código P (por vez primera) al inicio del siguiente subcuadro. La palabra HOW es un auxiliar de adquisición para el código P. La estructura del mensaje de datos de 30 bits se indica en la tabla 54-3.

El subcuadro 1 contiene el bloque de datos I, el cual a su vez contiene cuatro parámetros para hacer una corrección cuadrática al tiempo del reloj del satélite que se indica. Asimismo, en el bloque de datos I hay ocho parámetros para hacer una corrección aproximada considerando el retraso ionosférico para aquellos usuarios no equipados con receptores de doble frecuencia (L1-L2). El bloque de datos II ocupa los subcuadros 2 y 3 y contiene la efemérides completa y exacta del satélite que se recibe. El subcuadro 4 es para mensajes de transmisión especial. El subcuadro 5 contiene los almanaques rotatorios de todos los satélites.

Además de la estructura de las señales que se transmiten, es de interés el entorno de razón de señal sobre ruido en que se recibirán las señales. En el diseño o análisis del rendimiento de los enlaces de satélite a Tierra, es usual formular la razón de potencia de señal recibida dividida entre el valor de la densidad espectral de potencia de ruido blanco efectiva en el receptor. Ésta es la llamada razón  $C/N_0$ .

La potencia de señal recibida  $C$  es igual a la transmitida, multiplicada por diversos factores de ganancia y pérdida que afectan el enlace. La principal entre éstas es la pérdida de propagación o de trayectoria, que es simplemente la atenuación de potencia en la onda electromagnética transmitida debido a la dispersión esférica del frente de onda con la distancia desde la fuente. Este factor de pérdida,  $L_s$ , está dado por:

$$L_s = (V_c/4\pi FR)^2 \quad (54-45)$$

donde  $V_c$  = velocidad de la luz, m/s

$R$  = distancia, m

$F$  = frecuencia, Hz

Debe tenerse presente que las unidades de  $V_c$  y  $R$  deben ser compatibles.

Se observa que la pérdida de propagación varía inversamente al cuadrado de la frecuencia y de la distancia. Por tanto  $0 < L_s < 1$ , y  $L_s$  siempre hará menor la potencia recibida que la transmitida. Como  $L_s$  varía con la distancia, existe un valor máximo y otro mínimo de  $L_s$ , dependiendo del ángulo de elevación del satélite visto desde la posición del usuario, que se supone cer-

Tabla 54-3. Estructura del mensaje de datos de 30 bits

Subcuadro	Tipo de palabra		Estructura
1	TLM <sup>a</sup>	HOW <sup>b</sup>	Bloque de datos I (corrección del reloj)
2	TLM	HOW	Bloque de datos II ("efemérides")
3	TLM	HOW	Bloque de datos II, cont. ("efemérides")
4	TLM	HOW	Bloque de mensajes
5	TLM	HOW	Bloque de datos III ("almanaque")

<sup>a</sup> TLM, palabra de telemetría.<sup>b</sup> HOW, palabra de entrega.

cana a la superficie de la Tierra. La distancia máxima (a un ángulo de elevación 0°) y la distancia mínima, respectivamente, son:

$$\begin{aligned} R_{\text{máx}} &= 25\,231 \text{ km} \\ R_{\text{mín}} &= 19\,652 \text{ km} \end{aligned} \quad (54-46)$$

suponiendo una Tierra esférica con radio medio igual a 6 371 km. Por tanto, existen valores máximo y mínimo de  $L_s$  en las frecuencias L1 y L2, como se ilustra en la tabla 54-4. En esta tabla se dan los valores escalar y en decibels de las pérdidas, donde  $X \text{ dB} = 10 \log(X)$ .

La cantidad de ruido efectiva en el receptor es, para el GPS, esencialmente el ruido que se genera en el preamplificador de radiofrecuencia, que es de diseño especial para bajo ruido. Esto implica que el preamplificador de bajo ruido tiene la ganancia suficiente (20 a 30 dB) para superar el ruido generado en las primeras etapas del receptor mismo. En este caso, la densidad espectral de ruido blanco,  $N_0$ , está dada por:

$$N_0 = KT_s \quad (54-47)$$

donde  $T_s$  es la "temperatura de ruido del sistema" en kelvins.  $T_s$  es la temperatura de ruido del preamplificador mismo, más cualquier pérdida de cable y conector entre la antena y el preamplificador. Por lo general se hacen grandes esfuerzos para montar el preamplificador próximo a la antena. La constante  $K$  es la constante de Boltzmann, dada por:

$$K = 1.38 \times 10^{-23} \text{ W/Hz/K} \quad (54-48)$$

La temperatura de ruido del sistema  $T_s$  puede especificarse conforme al "factor de ruido" estándar  $F$  del sistema por medio de:

$$T_s = 290^\circ(F - 1) \quad (54-49)$$

Un medio útil para determinar los valores de  $C/N_0$  disponibles en los diversos casos es una tabla de control del diseño. En esta tabla se introducen los diversos parámetros del enlace, tales como las potencias transmitidas, ganancias, pérdidas, densidad espectral de ruido, etc., para llegar a un valor de  $C/N_0$  disponible. Este valor se emplea entonces para calcular niveles de rendimiento de diversas partes del receptor.

Una tabla de ejemplo para el canal C/A en L1 es la tabla 54-5. Con respecto a esta tabla se hacen los siguientes comentarios.

La pérdida de modulación explica el hecho de que la potencia total del transmisor del satélite se distribuye entre señales C/A y P en L1 y la señal en L2. Las proporciones de potencia son:

$$C/A (L1):P(L1):L2 = 4:2:1 \quad (54-50)$$

Las tolerancias explican las incertidumbres en las especificaciones, cambios con el envejecimiento o la temperatura, o variaciones en la relación geométrica entre el usuario y el satélite. Para un usuario con una antena nominalmente omnidireccional sobre el hemisferio superior, existe una variación grande en la ganancia entre la dirección del cenit y el horizonte. La tolerancia refleja esto. Asimismo, una antena omnidireccional que se polariza en forma circular en el cenit se vuelve elíptica a menores ángulos de elevación. De ahí la tolerancia para la pérdida por polarización.

El valor nominal de  $C/N_0$  en la tabla es 48.4 dB y la tolerancia es -12.9 dB. Por tanto, el valor disponible es +48.4 dB en el mejor de los casos y quizá tan reducido como +35.5 dB en el peor. La experiencia ha demostrado que las tolerancias adversas difícilmente llegan a sumarse en forma lineal, salvo aquellas que reflejan efectos correlacionados. Tres tolerancias correlacionadas de este tipo son la de la pérdida de propagación, la pérdida por polarización de la antena y la ganancia de la antena en la recepción. La suma de estas tres tolerancias correlacionadas es -8.5 dB, y se puede esperar que ocurra a los ángulos de elevación bajos, por ejemplo, menores de 10°.

La disponibilidad de 48.4 dB de  $C/N_0$  en el mejor de los casos fija el límite superior de las posibilidades de procesamiento del receptor en el ejemplo que se

Tabla 54-4. Pérdida de propagación de las portadoras L1 y L2

	L1	L2
$R_{\text{mín}}$	$5.9460 \times 10^{-19}$ (- 182.3 dB)	$9.7926 \times 10^{-19}$ (- 180.1 dB)
$R_{\text{máx}}$	$3.607 \times 10^{-19}$ (- 184.4 dB)	$5.9408 \times 10^{-19}$ (- 182.3 dB)

Tabla 54-5. Tabla de control de diseño, canal de libre acceso, frecuencia L1

Parámetro	Valor nominal (dB)	Tolerancia (dB)
1. Potencia del transmisor	13.0 dBW	0
2. Pérdida de modulación, división de potencia	- 2.4	0
3. Pérdida del circuito, transmisor	- 0.3	- 0.1
4. Ganancia de la antena, transmisión	+ 13.7	- 2.2
5. Pérdida de direccionamiento de la antena, transmisión	0	0
6. Pérdida de propagación; $F = 1575.4$ MHz, $R = 19\,652$ km	- 182.3	- 2.1
7. Pérdida de polarización de la antena	0	- 1.4
8. Ganancia de la antena, recepción (omni)	+ 3.0	- 5.0
9. Pérdida de direccionamiento de la antena, receptor	0	0
10. Pérdida del circuito, receptor	- 0.3	- 0.2
11. Pérdida del enlace neto (suman de 2 a 10)	- 168.6	- 11.0
12. Potencia total recibida, $C(1 + 11)$	- 155.6 dBW	- 11.0
13. Densidad espectral de ruido del sistema, $N_0$		
Factor del ruido del pream. 3.0 dB $\left\{ \begin{array}{l} + 1.0 \\ - 0.0 \end{array} \right.$		
Temperatura de pérdida + 20 K, nominal	- 204.0 dBW/Hz	+ 1.9
14. $C/N_0$ recibido (12 - 13)	+ 48.4	- 12.9

presenta. Aplicando la regla empírica según la cual se requieren razones de señal sobre ruido de 10 dB o mayores para lograr un procesamiento óptimo, se deduce que el mayor ancho de banda de procesamiento final no puede ser mayor de unos 7 kHz. Siendo más conservadores, en vista de las tolerancias adversas, este ancho de banda de procesamiento final podría reducirse por ejemplo a 3 kHz. Los anchos de banda reales que se utilizan para el procesamiento final dependerán, desde luego, de consideraciones más detalladas. Pero cuando menos el valor de  $C/N_0$  disponible fija la escala para hacer una consideración más a fondo.

#### 54.7.2 Técnica de telemetría porseudoruido (PN) del GPS

La técnica de telemetría PN del GPS es una descendente lineal directa del esquema original desarrollado por el Jet Propulsion Laboratory de la NASA durante la década de 1960<sup>12</sup>. Aunque la telemetría PN se consideró durante la década de 1950<sup>13</sup>, se perfeccionó en JPL y se aplicó en la mejor época de los vuelos espaciales a proyectos como Pioneer, Mariner, Surveyor, Viking, Apollo y Voyager. La misma se sigue empleando actualmente en los transbordadores espaciales y otros proyectos.

La idea básica subyacente a la telemetría PN es la que se indica a continuación. Intervienen dos códigos PN idénticos. Un código se propaga entre satélite y usuario. El otro se conserva en el equipo del usuario. Como un código PN es totalmente determinista, la serie de unos y ceros del código se conoce en su totalidad. Ambos códigos se inician en sincronía a un tiem-

po conocido específico,  $t_0$ . El código que se propaga entre el satélite y el usuario es recibido por este último, demorado el tiempo de propagación. El usuario desplaza entonces su propia versión local del código, llevando un registro del retardo inyectado en el "código local", hasta que su secuencia coincida exactamente con la que se recibe del satélite. Cuando el usuario observa que las dos versiones del mismo código PN se "acoplaron" o sincronizaron en el momento de observación  $t_r$ , anota que el tiempo de retardo de distancia  $\Delta T$  al momento  $t_r$  fue precisamente esa cantidad de tiempo que el usuario inyectó al código local para hacer que coincidiera con el código recibido.

El acoplamiento del código local con el recibido consiste en hacer coincidir las series de ceros y unos y después hacer coincidir "exactamente" los flancos de ascenso y de descenso de cada bit (o chip, como se denominan en telemetría) dentro de las series. Cualquier error en el acoplamiento de los códigos en el retardo se traduce en un error al determinar la distancia entre el satélite y el usuario. La conversión entre el error de retardo  $\Delta T$  y el de distancia  $\Delta R$ , mediante el uso de la velocidad de la luz,  $V_c = 2.99739 \times 10$  m/s, es:

$$\frac{\Delta R}{\Delta T} = 0.983 \text{ pies/ns} \quad (54-51)$$

A fin de conseguir que el problema del acoplamiento de códigos produzca mediciones de tiempo muy precisas, la duración de cada chip de código se hace muy pequeña. Para el código C/A, la duración del chip es de  $0.9775 \mu\text{s}$ . Para el código P, la duración es 97.75 ns.

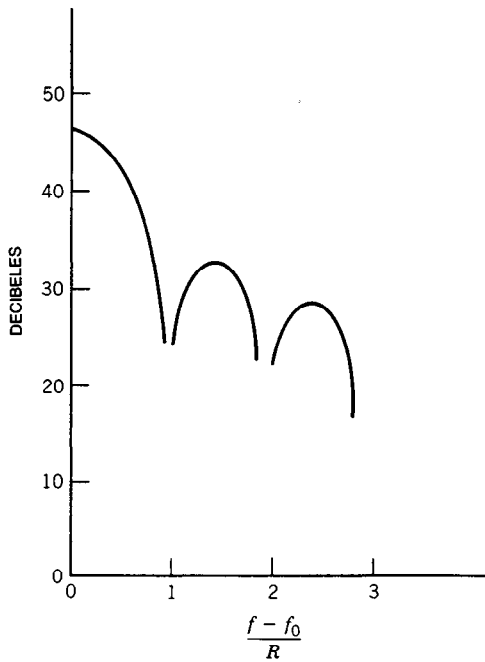


Fig. 54-6. Espectro de señales de pseudoruido (PN).

Debido al método que se aplica para hacer coincidir un chip de código con otro, la resolución en el tiempo de retardo es todavía menor que las duraciones mismas de los chips.

El uso de duraciones de chips de código por debajo de los microsegundos significa que el ancho de banda modulada de la señal telemétrica es del orden de los megahertz. En el caso de la manipulación por desfaseamiento de  $\pm 90^\circ$ , que se utiliza en el GPS para colocar la modulación del código en la portadora de radiofrecuencia, la señal modulada puede expresarse en la forma:

$$s(t) = AC(t)\cos(2\pi f_0 t) \quad (54-52)$$

donde  $A$  = amplitud de la señal

$f_0$  = frecuencia de la portadora, Hz

$C(t)$  = onda de forma de código analógica

La onda de forma de código analógica es  $+1$  o  $-1$ ; es decir, un chip 0 digital produce un  $+1$  analógico y un chip 1 digital produce un  $-1$  analógico. Como  $C(t)$  es característicamente una onda de forma rectangular, el espectro de potencia de frecuencia tiene la forma característica:

$$S(f) \sim \frac{\sin^2 \left[ \frac{\pi}{R} (f - f_0) \right]}{\left[ \frac{\pi}{R} (f - f_0) \right]} \quad (54-53)$$

donde  $f$  = frecuencia, Hz

$f_0$  = frecuencia de la portadora, Hz

$R$  = velocidad de transferencia de chips de código, chips/s

La gráfica de la ecuación 54-53 se presenta en la figura 54-6.

Se sabe que el 92% de la potencia en el espectro de señales PN reside entre los primeros puntos nulos en uno u otro extremo de la frecuencia portadora. Por tanto, en la mayor parte de los receptores se emplean anchos de banda de  $2R$  o mayores para transferir la señal modulada. Ahora bien, con un valor disponible de  $C/N_0$  de 48.4 dB/Hz, la razón de señal sobre ruido en el caso de la señal C/A en un ancho de banda de  $2R = 2.046$  MHz es  $-14.7$  dB. De este modo, los chips de señales no se pueden observar en tal ancho de banda, debido al efecto dominante del ruido acompañante.

En virtud de la imposibilidad de sincronizar los códigos recibido y local en una base de bit por bit, se aplica un método más directo, que puede operar en un ancho de banda menor. En este método se emplean las propiedades de "correlación" de los códigos PN.

Supóngase que por el momento se ignora el hecho de que el código PN recibido existe como modulación en una onda de forma portadora sinusoidal. Se considerará el problema más simple de sincronizar dos ondas de forma de código PN idénticas que existen en el formato analógico  $\pm 1$ , y se supondrá que el código recibido directamente se introduce en una puerta de un multiplicador analógico como  $C(t)$ . Sea el código local la entrada a la segunda puerta del multiplicador como  $C(t + \tau)$ , para denotar una ligera desincronización de  $\tau$  segundos. Sea la salida del multiplicador analógico procesada por un promediador de tiempo, tal como un filtro pasabajas. Esta operación se ilustra en la figura 54-7.

El dispositivo que se representa en la figura 54-7 se denomina correlacionador cruzado de promediación temporal. Enseguida se examina su operación con dos versiones relativamente demoradas del mismo código PN. Para esto considérese también la figura 54-8. En esta figura se muestran varios chips representativos del código PN directo y el código PN local cuando están casi en sincronía. Las dos gráficas de arriba son los códigos mismos, mientras que la de abajo es el producto de los códigos,  $C(t + \tau) \cdot C(t)$ .

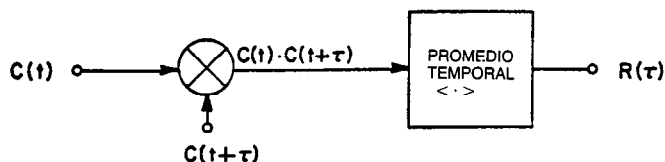


Fig. 54-7. Correlacionador cruzado.



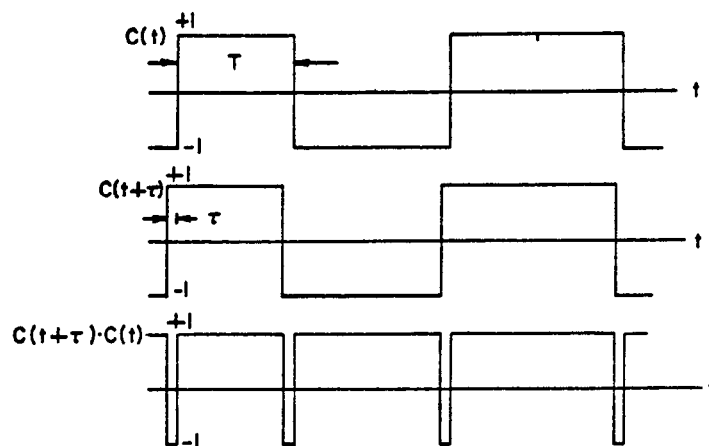


Fig. 54-8. Códigos de seudoruido (PN) con retardo relativo. Las dos gráficas de arriba muestran códigos PN directo y local; la gráfica de abajo muestra el producto de los códigos.

Es evidente, por inspección de la figura 54-8, que el producto  $C(t + \tau) \cdot C(t)$  es +1 la mayor parte del tiempo, con rápidas excursiones al estado -1 durante esos periodos breves de duración  $\tau$ , cuando  $C(t)$  y  $C(t + \tau)$  son de signo opuesto. También es claro que el promedio temporal del producto es positivo y casi +1. Cuando  $\tau = 0$ , o los códigos están exactamente sincronizados, el promedio es exactamente +1. Se puede demostrar que la salida o producto  $R(\tau)$  del promedio varía linealmente con la desviación o retardo relativo,  $\tau^{12}$ . Asimismo, para un código PN muy largo, el promedio es en esencia cero cuando  $C(t)$  y  $C(t + \tau)$  están desincronizados en más de un periodo de chip,  $T$ . Esta función,  $R(\tau)$ , que es la función autocorrelación del código PN básico, se ilustra en la figura 54-9.

Las propiedades de correlación de los códigos PN hacen posible la sincronización de los códigos PN recibido y local mediante el uso de un ancho de banda menor que el del código mismo. El menor ancho de banda es el del filtro pasabajas de promediación temporal, que se emplea en el intercorrelacionador. Por ejemplo, si se deseara observar  $R(\tau)$  con sólo el  $\pm 5\%$  de ruido con un valor disponible de  $C/N_0$  de 48.4 dB, se requeriría una razón de señal sobre ruido en la salida del promediador de 26 dB. Esto, a su vez, requeriría que el ancho de banda del filtro pasabajas de promediación fuera de 174 Hz.

Siguiendo con el ejemplo, otra pregunta interesante que surge es: si  $R(\tau) = 1$  significa la sincronización del código y se sabe que la salida del correlacionador tiene  $\pm 5\%$  del ruido, ¿cuál es el posible error de distancia en

que se incurre al aceptar cualquier medición individual de  $R(\tau)$  dentro de una vecindad del  $\pm 5\%$  de +1? A partir de la figura 54-9, con  $T = 293$  m equivalente, se puede determinar que la estrategia de telemetría podría incurrir en un error de  $\pm 29.3$  m.

Aquí se ha simplificado el tratamiento de la correlación de código PN para no perder de vista los resultados esenciales. En la práctica, la señal PN recibida puede existir como modulación PSK en una portadora sinusoidal, como en la ecuación 54-52. La multiplicación puede realizarse por el código local como modulación PSK en una sinusoide de frecuencia diferente. En este caso el multiplicador actúa como un mezclador para producir el producto de los códigos  $C(t) \cdot C(t + \tau)$  existente como modulación PSK en alguna sinusoide de frecuencia intermedia (IF). El filtro pasabanda IF realiza la promediación a fin de formar  $R(\tau)$ , que ahora está presente en la amplitud de la sinusoide de salida IF. Existen muchas formas de efectuar la correlación.

Las partes esenciales de la operación de telemetría PN son la forma de  $R(\tau)$  y el ajuste del retardo del código local para obtener una "estimación" de que  $R(\tau) = 1$ . El error en esta estimación se traduce en el error rms en la medición de la distancia. La velocidad con la cual se puede hacer una medición de distancia es directamente proporcional al tiempo que se requiere para encontrar el punto de sincronización  $R(\tau) = 1$ . La velocidad puede canjearse por exactitud de la estimación de la distancia en ciertos diseños de los receptores. En la sección que sigue se aborda el tema de las estrategias de búsqueda de la "correlación".

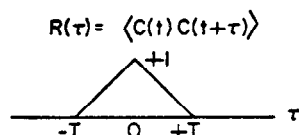


Fig. 54-9. Función autocorrelación de seudoruido (PN).

### 54.7.3 Técnicas de receptores en coherencia e incoherencia de fase

Existen varias posibilidades para formar  $R(\tau)$  y encontrar el retardo de código local que corresponde al centro del triángulo de la figura 54-9. Podría tratarse de

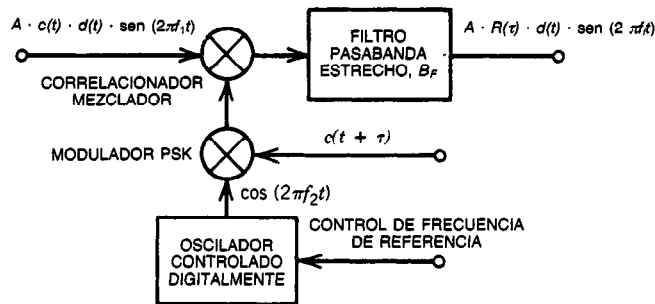


Fig. 54-10. Correlacionador de frecuencia intermedia (IF). PSK = manipulación por defasamiento.

implementar un servodispositivo de retroalimentación en el receptor que ajuste automáticamente el retardo de código local a fin de "seguir" el punto  $R(\tau) = 1$ . En caso contrario podría elegirse barrer el código local más allá del punto de sincronización con el código recibido, observar  $R(\tau)$  durante el barrido y determinar, después de éste, el retardo del código local que produjo  $R(\tau) = 1$ .

La elección entre los dos métodos de telemetría recién descritos depende de otros factores. Uno de éstos es si el receptor tiene "canales múltiples". Esto significa si se puede destinar en el receptor un canal analógico independiente a cada uno de los cuatro satélites que se requieren. Varios de los primeros receptores militares GPS se diseñaron de esta manera. En tal caso, pueden emplearse cuatro sistemas de retroalimentación independientes para impulsar cuatro generadores individuales de código local a fin de seguir de manera simultánea los cuatro códigos recibidos de distancia del satélite.

La alternativa a un receptor de canales múltiples que siga de manera simultánea cuatro códigos de distancia del satélite es un receptor "secuencial" que hace mediciones de distancia, en forma alternativa, entre los cuatro satélites. En este caso se conmuta secuencialmente un solo generador y correlacionador de código para sincronizarlo con los códigos de distancia del satélite que llegan, uno a la vez. Una unidad de demostración militar pionera basada en este concepto produjo un tiempo de conmutación mínimo entre los satélites de 1.2 s. No obstante, se puede hacer que un receptor secuencial tenga un desempeño similar al de un receptor de canales múltiples si la conmutación se puede efectuar lo suficientemente rápido. Este receptor emplea un alto nivel de procesamiento digital y se denomina receptor múltiplex<sup>14</sup>.

Considérese ahora la señal recibida tal y como podría existir en la salida de un correlacionador de código hipotético implementado por un mezclador en la formación de la última frecuencia intermedia del receptor. La figura 54-10 ilustra los circuitos. En esta figura, la señal de entrada al correlacionador mezclador tiene la frecuencia  $f_1$ . Un oscilador controlado digitalmente produce un coseno de frecuencia  $f_2$ , que es manipulado por defasamiento por el código local,  $C(t + \tau)$ . El mezclador forma el producto de códigos y el

filtro pasabanda lo promedia, al mismo tiempo que también selecciona la frecuencia IF  $f_i = f_1 - f_2$ . El ancho de banda del filtro pasabanda,  $B_f$ , es lo suficientemente pequeño para promediar el producto de los códigos, pero lo suficientemente grande para permitir el paso de la modulación de los datos,  $d(t)$ .

Ahora es posible procesar la salida del correlacionador en una de varias formas. Si se desea recuperar los datos, puede aplicarse la señal IF a la entrada de un ciclo de Costas<sup>15</sup> (fig. 54-11). El ciclo de Costas regenera un senoide de referencia que es coherente en fase con la frecuencia IF suprimida  $f_i$ . La referencia se aplica a un detector de amplitud sincrónico (detector de productos), que también es activado por la señal recibida. Se aplica una versión defasada  $90^\circ$  a un detector semejante. Los dos detectores desmodulan la señal IF y producen dos componentes de señales denominadas componente  $I$  (en fase) y componente  $Q$  (en cuadratura). Estas componentes se expresan como:

$$\begin{aligned} I(t) &= \frac{A}{2} R(\tau) d(t) \cos \phi \\ Q(t) &= -\frac{A}{2} R(\tau) d(t) \sin \phi \end{aligned} \quad (54-54)$$

En la ecuación 54-54,  $\phi$  es el error (seguimiento) en fase entre la senoide IF regenerada y la senoide IF recibida. Para la operación adecuada del ciclo,  $\phi = 0$  o  $\pi$  y el canal  $I$  contiene la onda de forma de datos, ponderada en amplitud por  $\pm R(\tau)$ . Obsérvese que si  $\phi = \pi$ , entonces los datos en el canal  $I$  se invierten. El valor absoluto de la onda de forma del canal  $I$  produce  $(A/2)|R(\tau)| = (A/2) R(\tau)$ , ya que  $d(t) = \pm 1$ . Por tanto, el canal  $I$  puede utilizarse para deducir tanto los datos como  $R(\tau)$ .

Un ciclo de Costas sincroniza y sigue la fase (y la frecuencia) de la senoide de señal IF y produce una salida proporcional a  $R(\tau)$ . Se pueden emplear dos ciclos de Costas en una configuración de ciclo de retroalimentación para sincronizar y seguir la frecuencia de reloj del código recibido. Esta configuración se denomina ciclo de reloj de adelanto y atraso. Este esquema se vale de dos mezcladores de correlación y dos generadores de código local (fig. 54-12).

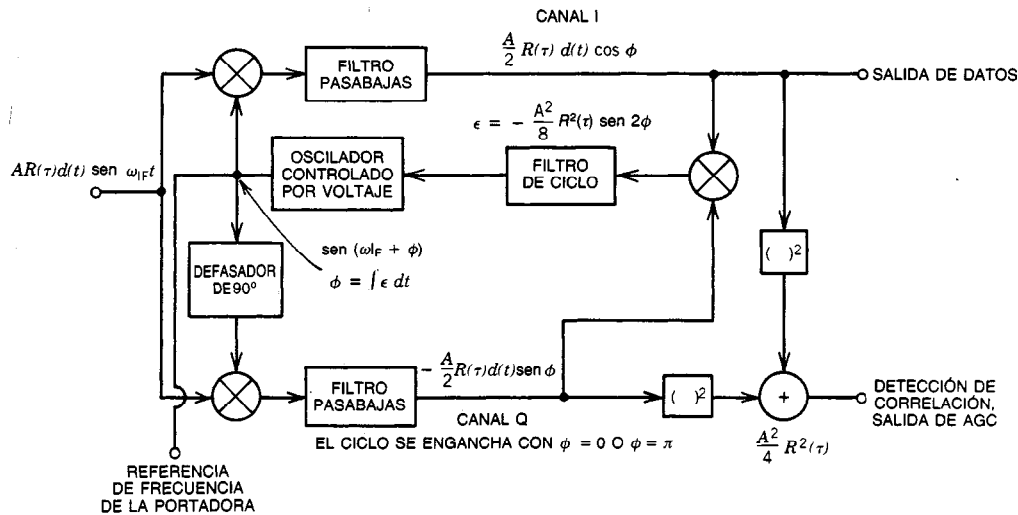


Fig. 54-11. Ciclo de Costas.

El ciclo de adelanto y atraso funciona de la manera siguiente. Los dos generadores de código local son accionados por el mismo generador (oscilador) de pulsos de reloj de código. Sin embargo, un generador de código se adelanta ligeramente respecto al otro. Por tanto el código de un generador está ligeramente "adelantado" respecto al código del otro generador (p. ej., 1 bit). Cuando ambos generadores se aproximan al punto de sincronización, el "código adelantado" lo sobrepasa y se "engancha" o sincroniza el "ciclo de Costas adelantado", lo que produce un " $R(\tau)$  adelantado", que se designa como  $R_E(\tau)$ . Cuando el código adelantado está 1/2 bit después del punto de sincronización, el "código retardado" está 1/2 bit antes de dicho punto; entonces el "ciclo de Costas retardado" se "engancha" o sincroniza y produce un " $R(\tau)$  retardado", que se designa como  $R_L(\tau)$ .

Volviendo a la figura 54-12, se advierte que  $R_L(\tau)$  se resta de  $R_E(\tau)$ , formando una "función error de seguimiento" de  $\tau$ . Esta función se ilustra en la figura 54-13. Cuando los códigos adelantado y atrasado están exactamente 1/2 bit adelantado y atrasado respecto al código recibido, entonces la señal de error de seguimiento es exactamente cero. Cuando el código recibido tiende a desplazarse en uno u otro sentido de la posición entre barras, la señal de error de seguimiento aumenta o disminuye desde cero y se utiliza para ajustar la frecuencia del generador de pulsos de reloj de código en la dirección adecuada, para conservar el "enganche del código". Cuando el ciclo se engancha en forma estable, ni el código adelantado ni el atrasado están exactamente sincronizados o "prestos". Cada uno está fuera de sincronización en 1/2 bit, pero el error se conoce con exactitud, de manera que se puede deducir un retardo  $R(\tau) = 1$ .

Es posible aplicar otro método de procesamiento cuando no se requieren datos. Éste podría ser el caso, por ejemplo, en un receptor secuencial de dos canales, donde un canal se puede destinar a obtener datos y el

otro sólo a telemetría. En el caso de sólo telemetría puede utilizarse un par de detectores  $I$  y  $Q$ , como en el ciclo de Costas, con un oscilador de referencia sinusoidal no enganchado o sincronizado en fase con la portadora recibida. Si este oscilador es diferente en frecuencia en una cantidad  $\Delta$  de la frecuencia  $IF$ , entonces las señales de los canales  $I$  y  $Q$  son:

$$I(t) = \frac{A}{2} R(\tau) d(t) \cos(2\pi\Delta t) \quad (54-55)$$

$$Q(t) = -\frac{A}{2} R(\tau) d(t) \sin(2\pi\Delta t)$$

Las dos señales de esta ecuación se pueden elevar al cuadrado y sumarse para producir una señal:

$$I^2(t) + Q^2(t) = \frac{A^2}{4} R^2(\tau) \quad (54-56)$$

Cuando el generador de código local en su barrido sobrepasa el punto de sincronización, la señal  $R^2(\tau)$  puede utilizarse para estimar el retardo  $R(\tau) = 1$ , ya que  $0 \leq R(\tau)$  y existe una relación uno a uno entre  $R(\tau)$  y  $R^2(\tau)$ . Dado que para aplicar este último método de procesamiento no es imperativo que haya enganche de fase o frecuencia a la sinusoide  $IF$ , el método se denomina incoherente. El método del ciclo de Costas se conoce como coherente.

Debido a los efectos Doppler en la señal recibida, los métodos de telemetría coherente e incoherente requieren una búsqueda, no sólo del retardo  $\tau$ , sino también de la frecuencia.

Volviendo a la figura 54-10, el oscilador controlado digitalmente (DCO) debe ajustarse para contrarrestar los efectos Doppler. Es necesario que la señal  $IF$  se centre exactamente en el ancho de banda de procesamiento más estrecho que se encuentre en el resto del receptor. Al adquirir la primera señal del satélite,

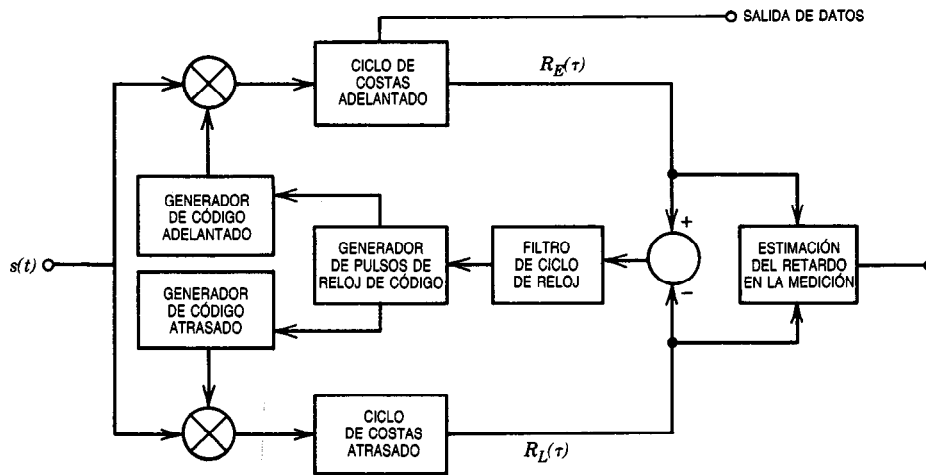


Fig. 54-12. Sincronizador de código de adelanto y atraso.

suele haber una incertidumbre de frecuencia considerable para buscar. Durante la navegación normal la incertidumbre suele ser mínima.

Para el procesamiento de mediciones coherente o incoherente, el DCO es precolocado y por consiguiente se busca la región de incertidumbre del retardo. Si no se observa evidencia de correlación, el DCO se re-coloca a una frecuencia diferente y se vuelve a buscar el retardo. De esta manera se busca la región bidimensional de incertidumbre en retardo y frecuencia.

En la tabla 54-6 se presenta un ejemplo hipotético del diseño de un sistema a fin de ilustrar algunos de estos aspectos. Este diseño es para un usuario aéreo (aeronave), con acceso exclusivo al código C/A, en la frecuencia L1, mediante una técnica de telemetría incoherente (de fase). Primero se hacen algunas suposiciones acerca de las exactitudes de velocidad y posición

que ofrece el filtro de navegación. Las desviaciones estándares del error de velocidad y el de posición se toman como  $3 \text{ m/s } \sigma_v$  y  $48.6 \text{ m } \sigma_p$ , respectivamente. A partir de  $\sigma_v$ , se calcula la incertidumbre resultante en la frecuencia recibida que se debe a un error en la velocidad del usuario. Se supone que la navegación es nominal y que los cálculos de la velocidad del satélite no contribuyen a la incertidumbre. El intervalo de búsqueda debido a la incertidumbre Doppler se toma, en forma conservadora, como  $\pm 2 \sigma_v$  o 63 Hz. Después, la incertidumbre de frecuencia a la IF debida a la inestabilidad del oscilador del usuario se calcula en 312 Hz. Por tanto, la incertidumbre de frecuencia total es 375 Hz. En teoría, el ancho de banda necesario para estimar el retardo  $R(\tau) = 1$ , aplicando la técnica incoherente, es de sólo 375 Hz. Sin embargo, como se utiliza un procesador digital de datos muestreados, el

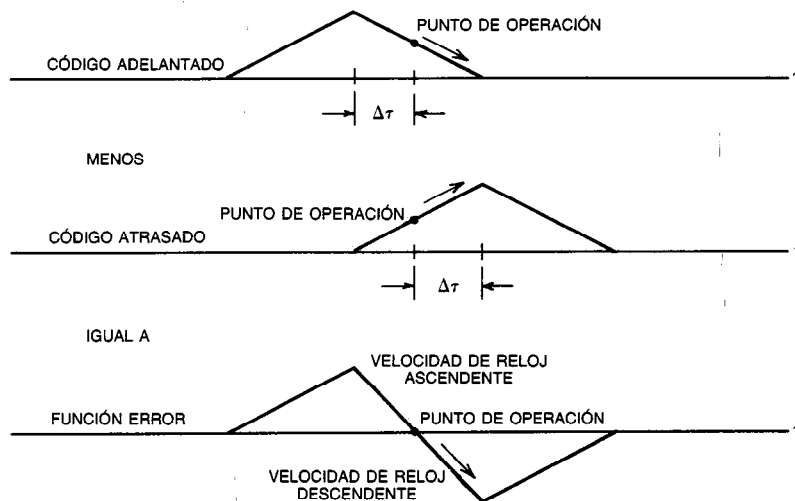


Fig. 54-13. Señal de error de seguimiento.

**Tabla 54-6. Ejemplo de diseño hipotético (aeronave, secuencial, L1, sólo libre acceso, medición incoherente)**

Parámetro	Valor
<b>Búsqueda Doppler</b>	
Incertidumbre de velocidad	$\sigma_v = 3 \text{ m/s}$
Distancia de búsqueda Doppler ( $\pm 2 \sigma_v$ )	63 Hz
Inestabilidad del oscilador ( $\pm \sigma_{IF}$ )	312 Hz
Incertidumbre de frecuencia total	375 Hz
Ancho de banda del filtro de medición	3 000 Hz
<b>Búsqueda de retardo</b>	
Incertidumbre de posición, esférica	$\sigma_p = 48.6 \text{ m}$
Ventana de búsqueda de retardo ( $\pm 3 \sigma_p$ )	293 m
Intervalo de incremento del código/muestras por barrido	19.5 m/16
Barridos totales/muestras totales	32/512
Tiempo muerto de recolocación de muestras	1 ms
Tiempo de medición del satélite	0.512 s
Tiempo de situación (cuatro satélites)	2.048 s
<b>Razón de señal sobre ruido (SNR) del estimador de la distancia</b>	
$C/N_0$ disponible (en el mejor de los casos)	48.4 dB/Hz
Ancho de banda requerido	34.8 dB Hz
SNR de medición disponible	13.6 dB
Exactitud de la medición	$\sigma_r = 10.8 \text{ m}$
<b>Presupuesto de error</b>	
Estimador de la distancia	10.8 m
Contribución del filtro GDOP-NAV ( $\text{GDOP}^a = 2.0$ )	5.8 m
Contribución ionosférica	20.0 m
Contribución de la efemérides del satélite	11.0 m
Contribución de la troposfera	1.0 m
Incertidumbre de posición total	$\sigma_p = 48.6 \text{ m}$

<sup>a</sup> GDOP, dilución geométrica de la precisión.

ancho de banda se toma como el valor mucho mayor de 3 kHz, para hacer posible un rápido muestreo. Tres veces el recíproco de este ancho de banda es el periodo de una época del código C/A.

A continuación, la ventana de búsqueda del retardo se toma como 293 m o el ancho de un chip, que es  $\pm 3 \sigma_p$  para la incertidumbre de la posición. La señal  $R(\tau)$  se barre en forma repetida 32 veces, colectando 16 muestras por barrido, para un total de 512 muestras de  $R(\tau)$ . La resolución del retardo es 19.5 m, pero el número grande de muestras que se toma reduce el error de cuantización a proporciones insignificantes. Con un tiempo muerto de recolocación de 1 ms por muestra, el tiempo para obtener una medición de la distancia del satélite es de 0.512 s. Una secuencia completa de cuatro satélites emplea 2.048 segundos.

Después se determina el rendimiento del estimador de la distancia. El mejor valor disponible de  $C/N_0$  se toma del ejemplo anterior como 48.4 dB/Hz. En un ancho de banda de 3 kHz, la razón de señal sobre ruido resultante es 13.6 dB. La exactitud de estimación resultante se determina por medio de simulación Monte Carlo de los algoritmos de procesamiento de distancias como  $\sigma_r = 10.8 \text{ m}$ .

Por último, se genera un presupuesto de errores para calcular la incertidumbre de la posición,  $\sigma_p$ , que se supuso antes. El uso de los algoritmos de navega-

ción suma 5.8 m al error del estimador de la distancia de 10.8 m. El retardo ionosférico diurno se considera un error extra de 20 m en la posición final. El error troposférico se toma como 1.0 m. El error de efemérides del satélite en el peor de los casos se supone de 11.0 m. Por tanto, la incertidumbre de la posición total de la suma es 48.6 m.

Se observa así que el diseño es un proceso iterativo. Dado que parte del rendimiento se puede determinar únicamente por medio de la simulación por computadora, el diseño se debe "afinar" hasta que sean consistentes todas las entradas. Obviamente, el diseño se optimiza en relación con un valor específico de  $C/N_0$ .

#### 54.7.4 Procesamiento de la navegación: linealización y dilución geométrica de la precisión (GDOP)

El procesador de la navegación del receptor del GPS es un componente importante. En ésta y las secciones que siguen se muestran algunos de los cálculos que deben efectuarse. En el procesador se incorpora una o más computadoras digitales de la variedad de microprocesador. Por estudios realizados<sup>16</sup> se ha demostrado que se necesita una longitud de palabra digital de 40 bits (precisión de 32 bits) para mantener el redondeo de operaciones en un valor bajo.

El objetivo del procesador de navegación consiste en aceptar cada medición de pseudodistancia tal y como se obtiene y aplicarla a la solución del vector de coordenadas del usuario, que consta de  $x, y, z$  y  $b$  (sesgo [desviación] del reloj) en el sistema ECEF. Al hacer esto, el procesador debe resolver las ecuaciones keplerianas para determinar el vector de posición de cada satélite que se recibe. El procesador determina asimismo qué satélites están visibles y, de ellos, qué cuatro satélites se deben elegir para tener la mejor configuración geométrica de navegación (geometría de navegación). El procesador suele incorporar un filtro que minimiza el promedio estadístico de la suma de los cuadrados de errores debidos al ruido del sistema y a maniobras dinámicas del usuario. Además, este filtro ajusta automáticamente sus parámetros para tomar en cuenta variaciones percibidas en el entorno de razón de señal sobre ruido y la geometría de navegación.

Volviendo a las ecuaciones 54-2 a 54-6, las mediciones en el GPS son pseudodistancias,  $R_i$ , para  $i = 1, 2, 3, 4$ . Se tiene:

$$R_i = f_i(x, y, z, b) =$$

$$= \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} + b \quad (54-57)$$

donde las coordenadas del  $i$ -ésimo satélite son  $x_i, y_i, z_i$ , las coordenadas del usuario son  $x, y, z$ , y el desvío del reloj del usuario en metros es  $b$ .

Como sucede en la ecuación 54-7, el problema se linealiza suponiendo una posición y un desvío del reloj del usuario  $\hat{x}, \hat{y}, \hat{z}, \hat{b}$  aproximadamente iguales a las cantidades reales y la  $\hat{R}_i$  correspondiente, como se calcula a partir de la ecuación 54-57. Entonces se realiza el desarrollo lineal de primer orden como sigue:

Volviendo a las ecuaciones 54-2 a 54-6, las mediciones en el GPS son pseudodistancias,  $R_i$ , para  $i = 1, 2, 3, 4$ . Se tiene:

$$R_i \cong \hat{R}_i + \left. \frac{\partial f_i(\cdot)}{\partial x} \right|_{\hat{x}} \cdot (x - \hat{x}) + \left. \frac{\partial f_i(\cdot)}{\partial y} \right|_{\hat{y}} \cdot (y - \hat{y}) + \left. \frac{\partial f_i(\cdot)}{\partial z} \right|_{\hat{z}} \cdot (z - \hat{z}) + \left. \frac{\partial f_i(\cdot)}{\partial b} \right|_{\hat{b}} \cdot (b - \hat{b}) \quad (54-58)$$

Las ecuaciones para cuatro satélites pueden expresarse en forma de matriz vectorial como:

$$\begin{bmatrix} R_1 - \hat{R}_1 \\ R_2 - \hat{R}_2 \\ R_3 - \hat{R}_3 \\ R_4 - \hat{R}_4 \end{bmatrix} = H \cdot \begin{bmatrix} x - \hat{x} \\ y - \hat{y} \\ z - \hat{z} \\ b - \hat{b} \end{bmatrix} \quad (54-59)$$

donde  $H$  es la matriz jacobiana cuadrada de derivadas parciales ordenadas, como se dan en la ecuación 54-58. Ahora bien, por ejemplo, se tiene:

$$\frac{\partial f_i(\cdot)}{\partial x} = \frac{(x - x_i)}{\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}} \quad (54-60)$$

$$i = 1, 2, 3, 4$$

$$\frac{\partial f_i(\cdot)}{\partial b} = 1 \quad i = 1, 2, 3, 4$$

y de este modo los elementos de la matriz  $H$  son cosenos directores de la distancia dirigida de la posición del usuario a la posición del satélite (fig. 54-14). Sean  $\alpha, \beta, \gamma$  los ángulos de  $R_i$ , la distancia dirigida, con respecto a los ejes  $x, y, z$ , respectivamente. Por tanto:

$$H = \begin{bmatrix} \cos \alpha_1 & \cos \beta_1 & \cos \gamma_1 & 1 \\ \cos \alpha_2 & \cos \beta_2 & \cos \gamma_2 & 1 \\ \cos \alpha_3 & \cos \beta_3 & \cos \gamma_3 & 1 \\ \cos \alpha_4 & \cos \beta_4 & \cos \gamma_4 & 1 \end{bmatrix} \quad (54-61)$$

Observando la estructura de la matriz  $H$ , se descubre que  $H$  no se podría invertir si el usuario fuese colineal con dos satélites o si los cuatro satélites estuviesen sobre la superficie de un cono.

Cuando  $H$  se puede invertir, es posible despejar el "vector" de coordenadas del usuario como:

$$\begin{bmatrix} x \\ y \\ z \\ b \end{bmatrix} = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \\ \hat{b} \end{bmatrix} + H^{-1} \cdot \left\{ \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} - \begin{bmatrix} \hat{R}_1 \\ \hat{R}_2 \\ \hat{R}_3 \\ \hat{R}_4 \end{bmatrix} \right\} \quad (54-62)$$

Es posible definir de manera formal vectores de posición y distancia y sus formas diferenciales como:

$$\mathbf{X} = \begin{bmatrix} x \\ y \\ z \\ b \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix}; \quad \begin{matrix} \delta \mathbf{x} = \mathbf{x} - \hat{\mathbf{x}} \\ \delta \mathbf{R} = \mathbf{R} - \hat{\mathbf{R}} \end{matrix} \quad (54-63)$$

Por tanto, de la ecuación 54-62 se tiene:

$$\delta \mathbf{x} = H^{-1} \delta \mathbf{R} \quad (54-64)$$

que es la forma diferencial de la solución linealizada en notación de matriz vectorial.

Supóngase ahora que se considera que el vector de distancia diferencial,  $\delta \mathbf{R}$ , está sujeto a algún error estadístico aleatorio (ruido). Supóngase que cada elemento de  $\delta \mathbf{R}$  está sujeto a un ruido aleatorio de desviación estándar  $\sigma_{\delta R}$  y que los cuatro ruidos son estadísticamente independientes. La validez de esta suposición depende de los mecanismos de generación de pseudodistancias,  $R_i$ , y sus estimaciones de linealización,  $\hat{R}_i$ .

Con estas suposiciones se puede formar la matriz de covarianza de la solución diferencial,  $\delta \mathbf{x}$ , como:

$$\text{var}\{\delta \mathbf{x}\} = \text{var}\{H^{-1} \delta \mathbf{R}\} = \sigma_{\delta R}^2 [H^T H]^{-1} \quad (54-65)$$

Ahora, la desviación estándar de la solución diferencial se define como:

$$\sigma_{\delta p} = \sqrt{\sigma_{\delta x}^2 + \sigma_{\delta y}^2 + \sigma_{\delta z}^2 + \sigma_{\delta b}^2} \quad (54-66)$$

Los elementos  $\sigma_{\delta x}$ ,  $\sigma_{\delta y}$ ,  $\sigma_{\delta z}$  y  $\sigma_{\delta b}$  son los elementos de la diagonal principal de la matriz de covarianza. Por tanto se tiene:

$$\sigma_{\delta p} = \sigma_{\delta R} \sqrt{\text{traza} [(H^T H)^{-1}]} \quad (54-67)$$

donde la traza ( $\cdot$ ) es el operador matricial que suma los elementos de la diagonal principal. La cantidad debajo del radical tiene un nombre importante en el GPS. Se denomina GDOP (de *geometric dilution of precision*, dilución geométrica de la precisión). Es la cantidad que relaciona el error en la pseudodistancia con el error en la solución de la posición, a través de la geometría, como se expresa en la matriz  $H$ . Con referencia a esta definición se tiene:

$$\sigma_{\delta p} = \text{GDOP} \cdot \sigma_{\delta R} \quad (54-68)$$

La cantidad GDOP incluye el error en el elemento de desviación del tiempo,  $b$ . Un navegante puede estar más interesado en el error de posición tridimensional, sin que le importe  $b$ . Éste es el llamado PDOP, que se obtiene como:

$$\sigma(\delta_x, \delta_y, \delta_z) = \text{PDOP} \cdot \sigma_{\delta R} \quad (54-69)$$

donde  $\text{PDOP} = \sqrt{\text{traza} \{D[H^T H]^{-1} D\}}$  y

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

El factor GDOP se utiliza para seleccionar los cuatro mejores satélites para el problema de navegación entre los que se tienen a la vista. El rendimiento de navegación óptimo se obtiene para el valor mínimo de GDOP, que siempre es positivo. El GDOP mínimo teórico ocurre para cuatro satélites, cada uno separado de los demás por un ángulo de  $120^\circ$ . Este valor es 1.62. Por desgracia, desde el punto de vista de un navegante en la superficie de la Tierra, tres de los satélites óptimos estarían debajo del horizonte.

Lo que resulta bastante sorprendente es que, si se agrega la restricción de la mínima visibilidad del satélite, no aumenta mucho la GDOP. En el caso de tres satélites con ángulos de elevación de  $0^\circ$ , separados en acimut  $120^\circ$ , y un cuarto satélite más o menos cercano en el cenit, la GDOP = 1.74.

Es evidentemente necesario elegir los satélites de modo que se minimice la GDOP, y ésta es labor del procesador de navegación. Surge la pregunta de cómo realizar la selección en la mejor forma con un mínimo de carga de trabajo de cálculo para el procesador. El método inicial de la "fuerza bruta" consistía en determinar el número de satélites visibles,  $n$ , y después obtener la GDOP para todos los conjuntos de cuatro satélites de  $n$  visibles. El número de tales operaciones de cálculo es:

$$N_{\text{cál}} = \binom{n}{4} = \frac{n!}{(n-4)!4!} \quad (54-70)$$

Por ejemplo, si  $n = 7$ , entonces  $N_{\text{cál}} = 35$ .

Resulta que existen más algoritmos prácticos de selección que siguen produciendo un valor adecuado de GDOP<sup>16</sup>. Por ejemplo, supóngase que el navegante puede elegir satélites en trayectoria, perpendicular a la trayectoria y cercano al cenit (tres ortogonales) y tomar el cuarto por encima del horizonte, en la dirección del vector resultante de los tres primeros. Para esta combinación, la GDOP = 1.67. Este algoritmo requiere  $4(n-1)$  operaciones de cálculo, que es 24 para siete satélites, y las operaciones necesarias no son tan complejas como las del método "óptimo". En la simulación por computadora de un caso práctico, la GDOP promedio en seis horas fue 2.64 para el algoritmo de selección óptimo y 2.92 para el algoritmo recién citado<sup>17</sup>.

#### 54.7.5 Filtrado de Kalman: estándar y adaptativo

Dadas cuatro mediciones de pseudodistancias simultáneas, cuatro posiciones de satélites determinadas al mismo tiempo, y el vector de coordenadas estimado requerido y el vector de pseudodistancia, la posición del navegante se puede calcular a partir de la ecuación 54-62 sobre una base de punto por punto.

Desde luego, existen dificultades prácticas para tener todas estas series de datos necesarios al mismo instante cada vez que se desea obtener una solución de posición. Pero estos detalles prácticos de cálculo no se abordan aquí, a fin de no oscurecer los principios básicos implicados.

El aspecto más importante, en este momento, es si la ecuación 54-62 produce resultados suficientemente exactos o si se necesita algún filtrado o suavizado (alisamiento) de las soluciones. La respuesta depende totalmente de la exactitud que se desee. En el ejemplo de diseño de C/A de la tabla 54-6, el uso de una solución no filtrada incrementaría la incertidumbre de la posición sólo en un 10% extra en el caso de una GDOP de 2.0. Sin embargo, el ejemplo estaba destinado al caso de razón de señal sobre ruido óptima y mejor que la GDOP promedio. La aplicación de la tolerancia de  $C/N_0$  en el peor de los casos de  $-12.9$  dB de la tabla 54-5 incrementaría la incertidumbre sólo de posición de C/A hasta la vecindad de los 650 m sin filtrado y para la GDOP promedio de 3.0. Este rendimiento probablemente sería inaceptable para la mayoría de los navegantes.

Un filtro navegacional es muy práctico para un vehículo que pasa mucho de su tiempo en movimiento uniforme (no acelerado). En este caso es posible diseñar un filtro que reduzca los efectos del ruido en la medición de la pseudodistancia casi sin límite. No obstante, cuando el vehículo está sujeto a aceleración y movimiento de orden superior, como en el viraje, la situación no es tan favorable. La salida del filtro no necesariamente seguirá una entrada de movimiento no uniforme sin error. Un filtro "óptimo" está diseñado para "partir la diferencia" que existe entre los erro-

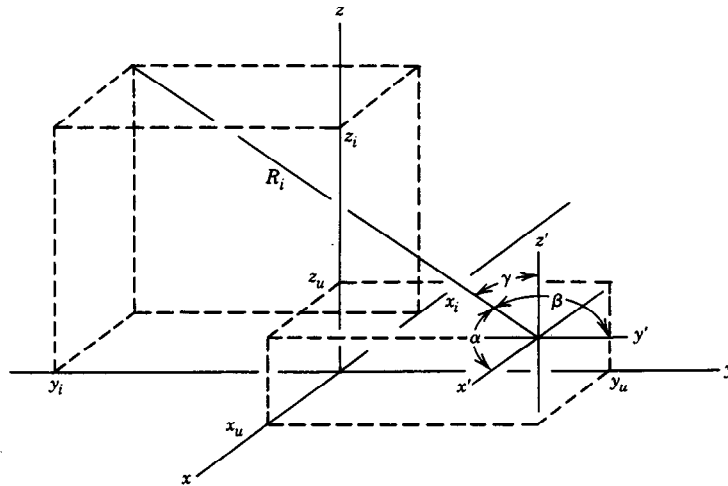


Fig. 54-14. Geometría relativa.

res debidos al ruido de la medición y los debidos al seguimiento deficiente por el filtro del movimiento no uniforme del vehículo. Es decir, un diseño de filtro óptimo intenta igualar en alguna medida los errores provocados por las dos fuentes diferentes. La medida suele ser el promedio estadístico del error total elevado al cuadrado.

Si un vehículo estuviese sujeto exclusivamente a movimientos cuyas gráficas pudieran ser descritas por polinomios de orden finito, en teoría sería posible crear un filtro para registrar este movimiento sin error. Por desgracia, un movimiento tan simple como una vuelta en círculo es descrito por sinusoides, cuyas derivadas superiores nunca terminan, y no por polinomios finitos. De este modo, los filtros sin errores resultan imposibles.

Un método popular para resolver el problema de filtrado consiste en describir el movimiento del vehículo como una función de muestreo a partir de alguna familia de funciones aleatorias. El ruido de la medición se describe de manera similar. En el caso en que el movimiento y el ruido del vehículo se pueden describir como funciones aleatorias gaussianas, existe un filtro lineal que minimiza, en todas estas funciones, el promedio estadístico del cuadrado del error de seguimiento total instantáneo. Dependiendo de ligeras diferencias, se denomina filtro de Kalman o filtro de Wiener. El filtro de Kalman es óptimo desde el momento en que se activa. Por el contrario, el filtro de Wiener, históricamente más antiguo, es óptimo sólo en el estado estable, después de que se ha extinguido el "transitorio de viraje".

Tanto los filtros polinomiales como los de Kalman requieren que la clase de entradas para la cual se espera que el filtro sea óptimo se modele en forma matemática. Por tanto, el diseño del filtro está basado en ese modelo de entrada específico. En el caso polinomial, si la entrada se puede describir como un polinomio de quinto orden, entonces el orden del filtro (número de polos) se hace proporcional para garan-

tizar una respuesta dinámica óptima a la función que se sigue. En el caso de Kalman, se especifica en realidad un filtro generador que produce la clase de entradas que se van a seguir. Los elementos del filtro generador terminan incorporándose a la estructura del filtro de Kalman.

El paso inicial para modelar la clase de entradas consiste en definir el "vector de estado" por seguir. Hasta aquí se ha definido el vector "de posición" del usuario como aquel que está compuesto de los elementos  $x$ ,  $y$ ,  $z$  y  $b$ . Pero esto no basta para los fines de la modelación. Para generar un modelo de entrada de orden superior que dé por resultado un filtro de orden superior, también se definirán derivadas respecto al tiempo de las componentes de la posición. En los receptores GPS es una práctica común modelar la posición, velocidad y aceleración para  $x$ ,  $y$ , y  $z$ , y modelar  $b$  y  $\dot{b}$  para la desviación del reloj. El motivo para limitar el modelo de  $b$  a sólo su primera derivada es que  $b$  proviene de un oscilador de reloj cuya frecuencia es inexacta. Existe poca aceleración en  $b$ . No se cumple lo mismo para  $x$ ,  $y$  y  $z$ .

El orden en el cual se insertan los diversos elementos en el vector de 11 estados no es arbitrario y tiene consecuencias de cálculo. Sin embargo, un orden específico se presta más fácilmente a una explicación. Las dos definiciones populares del vector de estado son:

$$\mathbf{x}^T = [x, y, z, \dot{x}, \dot{y}, \dot{z}, \ddot{x}, \ddot{y}, \ddot{z}, b, \dot{b}] \quad (54-71a)$$

$$\mathbf{x}^T = [x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}, z, \dot{z}, \ddot{z}, b, \dot{b}] \quad (54-71b)$$

El modelo generador de la matriz de vectores de  $x, y, z, b$ , para una u otra definición del vector de estado, se expresa en forma canónica de la manera siguiente:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{w}(t)$$

$$[x, y, z, b]^T = \mathbf{\Gamma}\mathbf{x}(t) \quad (54-72)$$





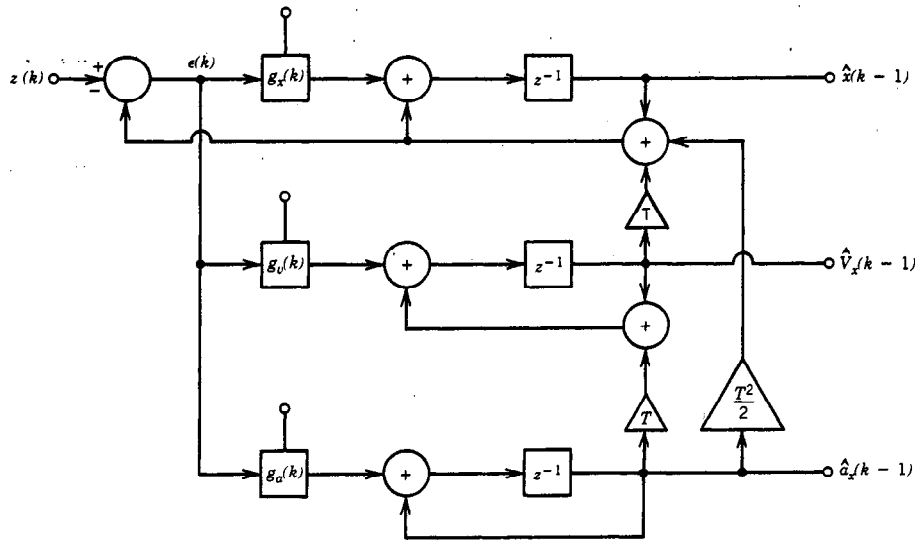


Fig. 54-15. Filtro de Kalman.

$$\Phi = \exp(AT) \quad (54-80)$$

$$\gamma = \int_0^T \exp(Aq) dq \cdot b$$

En la ecuación 54-80, la cantidad  $\exp(Aq)$  es la función exponencial de una matriz de  $3 \times 3$  y es en sí una matriz de  $3 \times 3$ .

Como la matriz  $A$  es una matriz de bloques de Jordan, la forma exponencial es:

$$\exp(AT) = \begin{bmatrix} e^{\lambda T} & T e^{\lambda T} & \frac{T^2}{2} e^{\lambda T} \\ 0 & e^{\lambda} & T e^{\lambda T} \\ 0 & 0 & e^{\lambda T} \end{bmatrix} \quad (54-81)$$

donde  $\lambda$  es el valor característico repetido único de la matriz  $A$ . Como  $\lambda = 0$  por inspección, se tiene:

$$\Phi = \exp(AT) = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \quad (54-82)$$

$$\gamma = \begin{bmatrix} T^3/6 \\ T^2/2 \\ T \end{bmatrix}$$

Ahora es posible expresar la ecuación 54-79 en forma de ecuación escalar acoplada como:

$$\begin{aligned} x(k+1) &= x(k) + T \cdot v(k) + T^2/2 \cdot a(k) + T^3/6 \cdot j(k) \\ v(k+1) &= v(k) + T \cdot a(k) + T^2/2 \cdot j(k) \\ a(k+1) &= a(k) + T \cdot j(k) \end{aligned} \quad (54-83)$$

Con las suposiciones de Kalman, la entrada de datos al filtro,  $z(k)$ , consta de coordenadas  $x(k)$  más ruido  $n(k)$ . Se supone que el ruido  $n(k)$  es independiente de  $j(k)$ , es cero, blanco y gaussiano y tiene media cero y varianza conocida  $\sigma_n^2$ . Así, la forma de los datos es:

$$z(k) = x(k) + n(k) \quad (54-84)$$

Las ecuaciones del filtro de Kalman se dan aquí junto con las ecuaciones generatrices y la forma de los datos, para hacer completa la exposición.

*Generatriz*

$$x(k+1) = \Phi x(k) + \gamma j(k)$$

$$x(k) = \lambda^T x(k); \lambda^T = [1, 0, 0]$$

*Datos*

$$z(k) = x(k) + n(k) \quad (54-85)$$

*Filtro*

$$\hat{x}(k) = \lambda^T \hat{x}(k)$$

$$\hat{x}(k) = g(k)[z(k) - \lambda^T \Phi \hat{x}(k-1)] + \Phi \hat{x}(k-1)$$

En la ecuación 54-85, la cantidad  $g(k)$  es un vector de ganancia, definido por:

$$\mathbf{g}^T(k) = [g_x(k), g_v(k), g_a(k)] \quad (54-86)$$

$\mathbf{g}(k)$  puede ser variante en el tiempo. El método para calcularla se describe con detalle como sigue.

Para comprender mejor la operación del filtro, las ecuaciones matriciales de vectores de la ecuación 54-85 se expresan en forma escalar acoplada como:

$$\begin{aligned} \hat{x}(k) &= \hat{x}(k-1) + T \cdot \hat{v}(k-1) + T^2/2 \cdot \hat{a}(k-1) + \\ &\quad + g_x(k) \cdot \varepsilon(k) \\ \hat{v}(k) &= \hat{v}(k-1) + T \cdot \hat{a}(k-1) + g_v(k) \cdot \varepsilon(k) \\ \hat{a}(k) &= \hat{a}(k-1) + g_a(k) \cdot \varepsilon(k) \\ \varepsilon(k) &= z(k) - [\hat{x}(k-1) + T\hat{v}(k-1) + \\ &\quad + T^2/2\hat{a}(k-1)] \end{aligned} \quad (54-87)$$

En la figura 54-15 se puede observar la estructura del filtro.

En la ecuación 54-87 y en la figura se advierte que cada nueva estimación consta de una predicción, basada en las estimaciones anteriores, más una corrección, basada en los nuevos datos. La corrección es un término de error ponderado en el cual el error son los datos menos la posición predicha. Los factores de ponderación son las "ganancias de Kalman." El término de error,  $\varepsilon(k)$ , se denomina residuo o innovación del filtro.

En el filtro estándar de Kalman, el vector de ganancia,  $\mathbf{g}(k)$ , se calcula a partir de un conjunto de ecuaciones matriciales de vectores acopladas que se exponen a continuación. Estas ecuaciones están basadas en la matriz de covarianza del error de filtrado, donde:

*Error de filtrado*

$$\tilde{\mathbf{x}}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k)$$

*Matriz de covarianza*

$$\mathbf{v}(k) = E[\tilde{\mathbf{x}}(k)\tilde{\mathbf{x}}^T(k)] \quad (54-88)$$

donde  $E[\cdot]$  es el promedio estadístico y bajo la suposición de que  $\mathbf{x}(k)$  y  $\hat{\mathbf{x}}(k)$  tienen ambos valores medios cero. Las ecuaciones de ganancia son:

$$\begin{aligned} \mathbf{v}(k|k-1) &= \Phi\mathbf{v}(k-1)\Phi^T + \sigma_f^2\gamma\gamma^T \\ \mathbf{g}(k) &= \mathbf{v}(k|k-1)\lambda[\sigma_n^2 + \lambda^T\mathbf{v}(k|k-1)\lambda]^{-1} \quad (54-89) \\ \mathbf{v}(k) &= [I - \mathbf{g}(k)\lambda^T]\mathbf{v}(k|k-1) \end{aligned}$$

En la ecuación 54-89,  $\mathbf{v}(k)$  es la matriz de covarianza calculada empleando la estimación predicha de un paso  $\hat{\mathbf{x}}(k|k-1)$ .  $\sigma_f^2$  y  $\sigma_n^2$  son las varianzas conocidas de la sacudida modelada y los procesos de ruido blanco aditivo, respectivamente.

El cálculo de  $\mathbf{g}(k)$  es recursivo y se inicializa con un valor  $\mathbf{v}(k=0)$ . Por lo general ésta es la matriz de varianza de estado estable del proceso,  $\mathbf{x}(k)$ . Sin embargo,

en el caso de navegación presente, con la estructura dada de la matriz  $\Phi$ , la varianza de  $\mathbf{x}(k)$  aumenta sin límite. Por tanto, se debe utilizar alguna otra inicialización adecuada para  $\mathbf{v}(k)$ .

Todos los elementos de ganancia de  $\mathbf{g}(k)$  tienden rápidamente a valores de estado estable. En dicho estado, el filtro de Kalman es exactamente el filtro de Wiener óptimo. Las ganancias de estado estable son función de  $T$ , el periodo de muestreo, y la razón  $\sigma_f^2/\sigma_n^2 = \text{SNR}$ , que es una razón de señal sobre ruido. Cuanto mayor es la razón, mayores son las ganancias, y viceversa. Cuanto mayores son las ganancias, mayor sensibilidad de respuesta tiene el filtro a cambios abruptos (maniobras) en la entrada, pero también es mayor el ruido blanco de medición que acepta el filtro. Cuanto menores son las ganancias, más lento se vuelve el filtro, pero menos ruidosa se vuelve la salida. En las figuras 54-16 y 54-17 se presentan ejemplos de la evolución de las ganancias y de la dependencia del estado estable respecto a la razón de señal sobre ruido.

Una dificultad que se observa con el uso de un filtro de Kalman en el problema de navegación es que la razón SNR cambia en forma drástica, dependiente de la maniobra del vehículo. Durante un movimiento no acelerado,  $\text{SNR} \rightarrow 0$  y las ganancias óptimas de estado estable son muy pequeñas, lo que da por resultado poco ruido en la salida del filtro. Durante un movimiento acelerado, la SNR aumenta y las ganancias óptimas de estado estable son mayores, lo que da por resultado una mayor aceptación de ruido blanco pero un error de filtro de dinámica inferior.

Por ejemplo, una aeronave con velocidad de 300 nudos sujeta a excursiones de control normales de  $\pm 3$  m con un periodo de 30 segundos y ruido en la medición de 10.8 m, tiene una  $\text{SNR} = 6.50 \times 10^{-6}$ . Durante una vuelta a velocidad normal de dos minutos,  $\text{SNR} = 1.54 \times 10^{-3}$ , una variación de +30 dB. En el cálculo estándar de la ganancia de Kalman no se tiene conocimiento de estos entornos radicalmente diferentes para la operación del filtro. Por tanto, para utilizar el filtro en forma efectiva, se debe emplear un algoritmo de ganancia "adaptativa".

No existe una teoría bien desarrollada de los filtros de ganancia adaptativa comparable a la teoría de Kalman y Wiener. Por tanto, los desarrollos en esta área son todos para un fin determinado y para casos especiales. El único requisito básico es que el residuo del filtro,  $\varepsilon(k)$ , se utilice para obtener información a fin de ajustar las ganancias del filtro. Las ganancias pueden ajustarse en forma independiente. Deben ajustarse en forma consistente con resultados producidos por la solución de las ecuaciones de ganancia. En caso contrario se puede generar una inestabilidad del filtro.

#### 54.7.6 Filtrado de Kalman extendido para el GPS

Existe una dificultad final en la aplicación del filtro de Kalman al problema de navegación del GPS: las mediciones de navegación ruidosas,  $R_i(k)$ , no constituyen la forma de datos correcta para el filtro de Kalman. El filtro requiere mediciones ruidosas de las coordenadas de posición. Las mediciones de pseudodis-

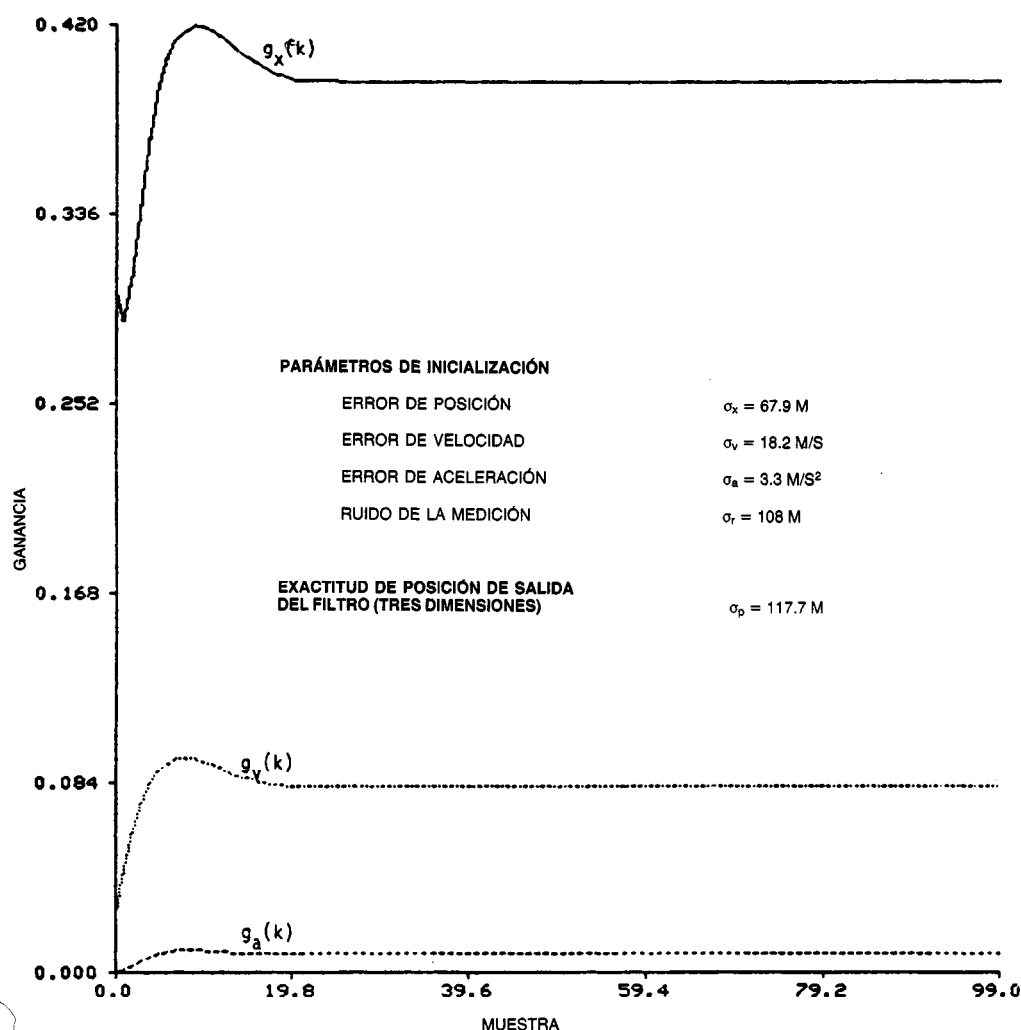


Fig. 54-16. Evolución de ganancias. Tomado de "The NAVSTAR Global Positioning System — A User's Approach to Understanding", curso breve de Altair Corporation, College Station, Texas, con permiso.

tancia, sin embargo, son transformaciones no lineales de las coordenadas de posición. Un método para resolver este problema consiste en utilizar la linealización que se desarrolló antes en este capítulo y que se incorporó en la ecuación 54-64. Esa relación se desarrolla aquí como:

$$\mathbf{x}(k) - \hat{\mathbf{x}}(k) = H^{-1}[\mathbf{R}(k) - \hat{\mathbf{R}}(k)] \quad (54-90)$$

donde  $\mathbf{x}(k)$  = vector de posición real

$\mathbf{x}^T(k) = [x, y, z, b]$

$\hat{\mathbf{x}}(k)$  = posición supuesta para los fines del cálculo de  $H^{-1}$

$\mathbf{R}(k)$  = vector de cuatro pseudodistancias medidas

$\hat{\mathbf{R}}(k)$  = predicción de  $\mathbf{R}(k)$  calculada mediante el uso de  $\hat{\mathbf{x}}(k)$  y la función no lineal de la ecuación 54-57.

Puesto que  $\mathbf{R}(k)$  tiene ruido blanco aditivo, la ecuación 54-90 agregará ruido blanco a  $\mathbf{x}(k)$  a través de la transformación lineal  $H^{-1}$ . Sin embargo, esta ecuación produce una transformación diferencial de  $\mathbf{R}$  a  $\mathbf{x}$ , y no una transformación global. De este modo persiste la pregunta: ¿cómo se utiliza la ecuación 54-90 para producir los datos de entrada para los filtros de Kalman?

Supóngase que en la ecuación 54-90, con el número de muestra  $k$ , el vector de posición supuesto  $\hat{\mathbf{x}}(k)$  está constituido por las componentes de posición predichas del filtro de Kalman, que tienen todas la misma forma que la componente  $x$ , dada como:

$$\hat{\mathbf{x}}(k|k-1) = \lambda^T \Phi \hat{\mathbf{x}}(k-1) \quad (54-91)$$

Ahora es posible reexpresar la ecuación 54-90, mostrando en forma explícita la contribución de ruido aditivo debido a  $H^{-1}\mathbf{R}(k)$ , como:

$$\begin{aligned} \mathbf{x}(k) + \mathbf{n}(k) - \hat{\mathbf{x}}(k|k-1) &= H^{-1}[\mathbf{R}(k) - \hat{\mathbf{R}}(k|k-1)] \\ &= \mathbf{z}(k) - \hat{\mathbf{z}}(k|k-1) = H^{-1}[\mathbf{R}(k) - \hat{\mathbf{R}}(k|k-1)] \\ &= \mathbf{e}(k) = H^{-1}[\mathbf{R}(k) - \hat{\mathbf{R}}(k|k-1)] \end{aligned} \quad (54-92)$$

Esto es, la cantidad  $H^{-1}[\mathbf{R}(k) - \hat{\mathbf{R}}(k|k-1)]$  puede identificarse como el término de error de cuatro vectores en un filtro de Kalman de 11 estados.

Debido a la elección que se hizo aquí de la matriz de acoplamiento  $\Phi$  en el modelo generador, ese modelo se reduce de uno de 11 estados a cuatro modelos independientes que tienen 3, 3, 3 y 2 estados, respectivamente, con base en la descripción de vectores de estado de la ecuación 54-71b.

El filtro de Kalman se puede componer, en forma semejante, de cuatro secciones independientes, cada una accionada por uno de los cuatro elementos del vector de error,  $\mathbf{e}(k)$ . El filtro de Kalman extendido, así entendido, se representa en forma de diagrama en la figura 54-18.

Aunque la estructura del filtro de Kalman se desacopla en forma óptima hasta donde concierne a la operación de cálculo, las estimaciones de la posición no son independientes, ya que los errores en  $x$ ,  $y$ ,  $z$  y  $b$  son acoplados en forma cruzada por  $\mathbf{f}(\cdot)$  y  $H^{-1}$ . Otro efecto es que las varianzas de ruido blanco en las entradas a los filtros  $x$ ,  $y$ ,  $z$  y  $b$  no son necesariamente iguales o fijas, ya que dependen de  $H^{-1}$ , que es variable con la geometría:

$$H^{-1} = \begin{bmatrix} l_{x1} & l_{x2} & l_{x3} & l_{x4} \\ l_{y1} & l_{y2} & l_{y3} & l_{y4} \\ l_{z1} & l_{z2} & l_{z3} & l_{z4} \\ l_{b1} & l_{b2} & l_{b3} & l_{b4} \end{bmatrix} \quad (54-93)$$

Sea el ruido aditivo en  $\mathbf{R}(k)$  indicado en forma explícita por:

$$\mathbf{R}(k) = \begin{bmatrix} R_1(k) \\ R_2(k) \\ R_3(k) \\ R_4(k) \end{bmatrix} + \begin{bmatrix} n_1(k) \\ n_2(k) \\ n_3(k) \\ n_4(k) \end{bmatrix} \quad (54-94)$$

Por tanto, los ruidos efectivos en  $x$ ,  $y$ ,  $z$  y  $b$  están dados por:

$$\begin{aligned} n_x(k) &= \sum_{i=1}^4 l_{xi} n_i(k) \\ &\vdots \\ n_b(k) &= \sum_{i=1}^4 l_{bi} n_i(k) \end{aligned} \quad (54-95)$$

Se supone que los ruidos de medición de la distancia  $n_i(k)$  tienen media cero, son independientes y presentan la misma varianza  $\sigma_n^2$ . Por tanto, las varianzas de  $n_x(k)$ ,  $n_y(k)$ ,  $n_z(k)$  y  $n_b(k)$  son:

$$\begin{aligned} \sigma_{n_x}^2 &= \sigma_n^2 \cdot \sum_{i=1}^4 l_{xi}^2 \\ &\vdots \\ \sigma_{n_b}^2 &= \sigma_n^2 \cdot \sum_{i=1}^4 l_{bi}^2 \end{aligned} \quad (54-96)$$

En este desarrollo del filtro de Kalman extendido se ha supuesto que las cuatro mediciones de pseudodistancias están disponibles al mismo tiempo y que el filtro total se opera en un ciclo con cuatro pseudodistancias disponibles. Éste no es el caso de un receptor secuencial. Se examinará la ecuación 54-91 con más detalle:

$$\begin{aligned} \begin{bmatrix} x - \hat{x} \\ y - \hat{y} \\ z - \hat{z} \\ b - \hat{b} \end{bmatrix} &= \begin{bmatrix} l_{x1} l_{x2} l_{x3} l_{x4} \\ l_{y1} l_{y2} l_{y3} l_{y4} \\ l_{z1} l_{z2} l_{z3} l_{z4} \\ l_{b1} l_{b2} l_{b3} l_{b4} \end{bmatrix} \cdot \begin{bmatrix} R_1 - \hat{R}_1 \\ R_2 - \hat{R}_2 \\ R_3 - \hat{R}_3 \\ R_4 - \hat{R}_4 \end{bmatrix} = \\ &= \begin{bmatrix} l_{x1}(R_1 - \hat{R}_1) \\ l_{y1}(R_1 - \hat{R}_1) \\ l_{z1}(R_1 - \hat{R}_1) \\ l_{b1}(R_1 - \hat{R}_1) \end{bmatrix} + \dots + \begin{bmatrix} l_{x4}(R_4 - \hat{R}_4) \\ l_{y4}(R_4 - \hat{R}_4) \\ l_{z4}(R_4 - \hat{R}_4) \\ l_{b4}(R_4 - \hat{R}_4) \end{bmatrix} \end{aligned} \quad (54-97)$$

La ecuación 54-97 muestra que el coeficiente  $l_{x1}$  introduce en el filtro la componente de distancia diferencial  $(R_1 - \hat{R}_1)$  para corregir la coordenada  $x$ . De modo similar,  $l_{y1}$  utiliza  $(R_1 - \hat{R}_1)$  para corregir la coordenada  $y$ , etc. La primera columna de  $H^{-1}$  corrige entonces las cuatro coordenadas de posición a partir de la información que proporciona la medición de la distancia diferencial  $(R_1 - \hat{R}_1)$ . De la misma manera la segunda, tercera y cuarta columnas de  $H^{-1}$  aplican las correcciones de los vectores de posición debidas al segundo, tercero y cuarto satélites, respectivamente.

Una pregunta interesante es, "¿qué sucede si estas correcciones se aplican en forma secuencial y no en forma simultánea?". Es decir, ¿qué sucede si el filtro opera en un ciclo cada vez que se obtiene una nueva medición de la distancia? La respuesta es que un procedimiento de este tipo funciona bien, siempre que la velocidad de muestreo sea alta en comparación con la velocidad de maniobra del vehículo. Por ejemplo, la figura 54-19 ilustra la salida real de un filtro de Kalman corregido en forma secuencial que opera con el GPS (sólo C/A) a bordo de un barco. La embarcación se dirige inicialmente al este a 10 nudos (5.16 m/s). Después vira lentamente al sur, ejecutando la vuelta de 90° en 55 segundos. La frecuencia de muestreo de la

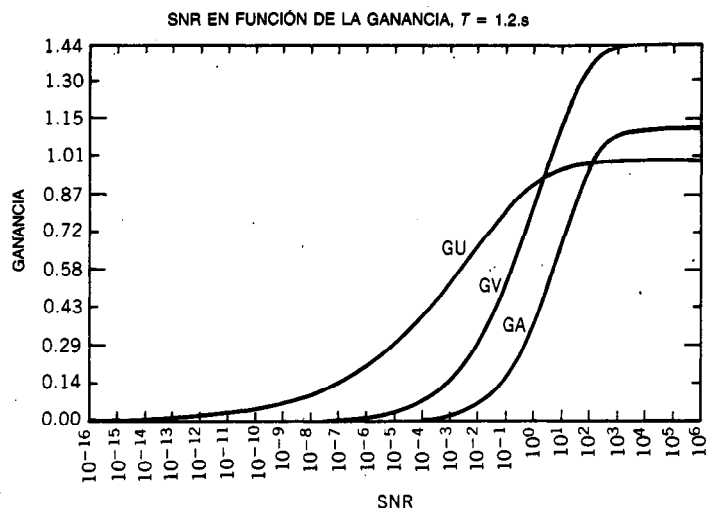


Fig. 54-17. Dependencia de las ganancias en función de la razón de señal sobre ruido (SNR). Tomado de "The NAVSTAR Global Positioning System - A User's Approach to Understanding", curso breve de Altair Corporation, College Station, Texas, con permiso.

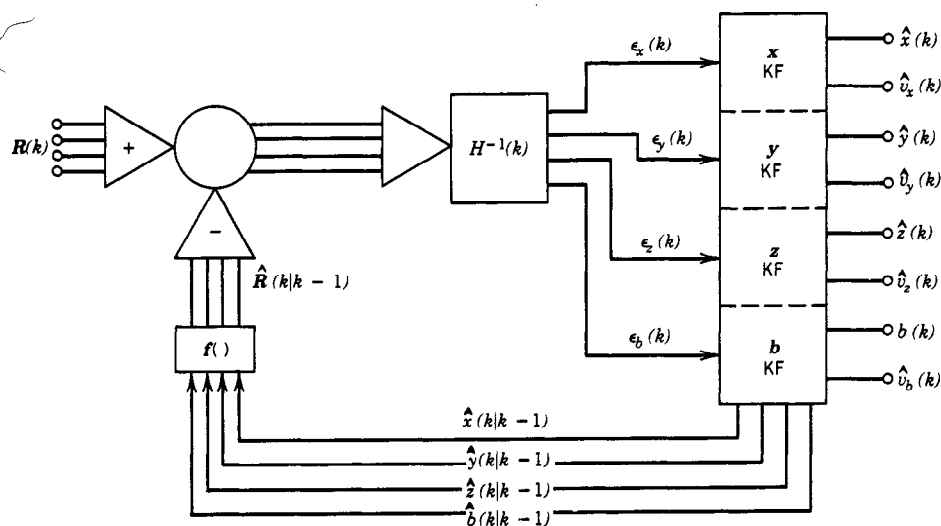


Fig. 54-18. Filtro de Kalman extendido (KF). Tomado de "The NAVSTAR Global Positioning System - A User's Approach to Understanding", curso breve de Altair Corporation, College Station, Texas, con permiso.

distancia es una medición cada 1.2 segundos. El seguimiento presenta poco ruido aleatorio, más o menos 6 m,  $1 \sigma$ , y no hay sobrecorrección apreciable en la vuelta. La corrección secuencial del filtro no tiene un efecto apreciable en este caso.

#### 54.7.7 Rendimiento del GPS y clasificación de errores

En el momento de escribir esto no se había publicado mucho sobre el rendimiento real del GPS. Lo poco que habían entregado los militares o llegado a impresión

acerca de experimentos civiles demostraba que el rendimiento era extraordinario. En realidad, la exactitud de localización geodésica del sistema es tan grande que no existe un patrón contra el cual se pueda medir. Parece probable que el GPS se convertirá en la norma de la medición de exactitudes geodésicas y de navegación.

El artículo publicado por Henderson y Strada<sup>17</sup> fue la primera versión "oficial" de los datos de prueba del GPS. En diversos simposios técnicos han figurado otros datos ilustrativos. El autor, por ejemplo, ha sido coinvestigador en un experimento marítimo civil con el GPS, del cual se han obtenido datos interesantes.

LOS LÍMITES DEL TIEMPO SON 3.34 A 3.37 GMT

FECHA DE CREACIÓN = 80194

LA POSICIÓN MEDIA REFERIDA AL TU ES  
28.42383 POR - 78.65839 GRADOS

##### SÓLO DATOS TU #####

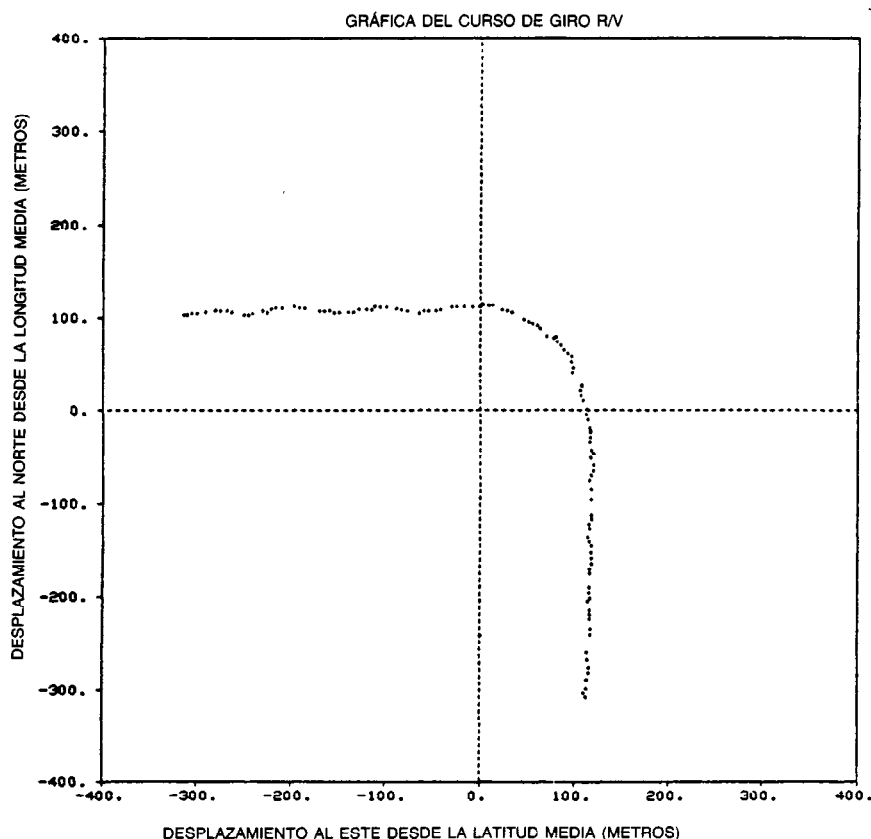


Fig. 54-19. Seguimiento del receptor secuencial.

El primer resultado de rendimiento verdaderamente informativo fue que el desempeño sólo de C/A no es 10 veces peor que el del código P, a pesar del factor de diferencia de 10 en la resolución del código. Si la frecuencia L2 se llegara a estandarizar para el código C/A, de manera que un usuario sólo de C/A pudiera tener la corrección ionosférica, es probable que los rendimientos de los códigos P y C/A fueran proporcionales para algunos usuarios.

Para los navegantes de superficie, las perturbaciones ionosféricas no son particularmente problemáticas, siempre que se disponga de tres satélites con buena dispersión en acimut y ángulos de elevación razonablemente grandes. En este caso, los errores ionosféricos horizontales tienden a cancelarse, dejando la contribución principal en la altitud. Con 18 satélites en operación, un navegante de superficie puede construir un algoritmo de selección de satélites para elegir satélites como se mencionó, a fin de minimizar los efectos ionosféricos horizontales.

Los datos sobre rendimiento de la referencia 17 (pág. 82) demostraron exactitudes de navegación absolutas del código P de 2.84 m de desviación y ruido de 5.47 m ( $1 \sigma$ ) en la navegación de superficie a bajas velocidades en el caso de un equipo de usuario. Otro equipo de código P en navegación de superficie a baja velocidad presentó una desviación de 2.84 m y ruido de 6.4 m ( $1 \sigma$ ) absolutos. Pruebas de vuelo de helicópteros indicaron rendimientos del código P de 9.6 m de desviación y ruido de 4.7 m ( $1 \sigma$ ) en tres dimensiones. Todos estos resultados del código P se tomaron en breves periodos de tiempo, de menos de 20 minutos.

En la referencia 17 (pág. 82) se presenta un resultado sólo de código C/A para un periodo de tiempo de seis minutos. En este caso se demostró una desviación de 30.3 m y ruido de 9.2 m ( $1 \sigma$ ) para la navegación en tres dimensiones de una aeronave de velocidad media. De la desviación de 30 m, 29.0 m se refieren a altitud. La exactitud horizontal de ese desempeño de vuelo fue 8.6 m de desviación y 8.3 m de ruido ( $1 \sigma$ ), absolutos.

Resultados sólo de código C/A para un barco atracado en Galveston, Texas, o Miami han demostrado en forma consistente desviaciones de menos de 30 m horizontales y ruidos de 6 m ( $1\sigma$ ) absolutos, a intervalos de dos horas. Las desviaciones se basan en los mejores mapas de puerto de pequeña escala disponibles. Estos últimos resultados de C/A fueron casos de mejor desempeño.

Los resultados de rendimiento del GPS calibrados cuidadosamente son escasos, aunque muchas evaluaciones del GPS se hallan en proceso. Es muy probable que los nuevos resultados de rendimiento seguirán demostrando la extrema exactitud del GPS, tanto con el código P como sólo con el código C/A.

#### REFERENCIAS BIBLIOGRÁFICAS

1. P. R. Escobal, *Methods of Orbit Determination*, Wiley, Nueva York.
2. R. H. Battin, *Astronautical Guidance*, McGraw-Hill, Nueva York.
3. T. A. Stansell, *The Transit Navigation Satellite System*, Magnavox Report R-5933, Torrance, CA.
4. T. O. Seppelin, "The Department of Defense World Geodetic System 1972", Defense Mapping Agency, Washington, DC.
5. K. W. Yip y O. H. Von Roos, "A New Global Ionospheric Model", *JPL Deep Space Network*, Jet Propulsion Laboratory, Pasadena, CA, pp. 70-87.
6. M. D. Eggers y D. M. Bullock, *Computer Simulation of Ionospheric Wave Propagation for Detection of Range Error for Satellite Navigation Systems*, Memorandum #81-06, Laboratorio de Telecomunicaciones y Sistemas de Control, Dpto. de Ingeniería Eléctrica, Texas A & M University, College Station, Texas, p. 44.
7. W. H. Guier y G. C. Weiffenbach, "A Satellite Doppler Navigation System", *Proc. IRE*, pp. 507-516.
8. J. H. Painter, "Designing Pseudo-Random Coded Ranging Systems", *IEEE Trans. Aero. Electr. Syst.* **AES-3**(1), pp. 14-27.
9. A. Bartholomew, "Satellite Frequency Standards", en *Global Positioning System*, The Institute of Navigation, Washington, DC, pp. 21-28.
10. S. W. Golomb, *Shift Register Sequences*, Holden-Day, San Francisco, p. 33.
11. J. J. Spilker, "Signal Structure and Performance Characteristics", en *Global Positioning System*, Institute of Navigation, Washington, DC, pp. 29-54.
12. S. W. Golomb, L. D. Baumert, M. F. Easterling y colaboradores, *Digital Communications With Space Applications*, Prentice-Hall, Englewood Cliffs, NJ, pp. 85-105.
13. H. W. Farris, M. P. Ristenbatt y colaboradores, *An Introduction to Pseudo-Random Systems*, Vol. 1, *Basic Concepts and Techniques*, Informe Técnico 104-1, Cooley Electronics Laboratory, University of Michigan, Ann Arbor.
14. P. Ward, "An Advanced NAVSTAR GPS Multiplex Receiver", *Record of IEEE 1980 Position, Location, and Navigation Symposium*, Atlantic City, NJ, pp. 51-58.
15. W. C. Lindsey y M. K. Simon, *Telecommunication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, pp. 62-64.
16. H. Parsiani, *A Navigation Algorithm for Single Channel Low-Cost GPS Receiver*, Dpto. de Ingeniería Eléctrica, Texas A & M University, pp. 28-45.
17. D. W. Henderson y J. A. Strada, "NAVSTAR Field Test Results", *Global Positioning System*, The Institute of Navigation, Washington, DC, pp. 234-246.





DECIMOSEGUNDA PARTE

---

# **Computadoras**

---



## CAPÍTULO 55

# Sistemas numéricos y representación de datos

**Alan B. Marcovitz**

Florida Atlantic University  
Boca Raton, Florida

### INTRODUCCIÓN

**55.1** Números binarios, octales y hexadecimales

**55.2** Suma binaria de números positivos

**55.3** Números con signo

**55.4** Aritmética binaria

**55.5** Decimal codificado en binario (BCD)

**55.6** Aritmética de precisión múltiple

**55.7** Aritmética de punto flotante

**55.8** Detección y corrección de errores

**55.9** Representaciones de datos alfanuméricos

---

## INTRODUCCIÓN

Este capítulo se dedica a la representación de información numérica y alfanumérica en una computadora digital. En la computadora, la información se codifica en palabras binarias de longitud estándar que constan de unos y ceros. La atención se concentra en el almacenamiento de una amplia gama de números y la aritmética relacionada con ellos. En el caso de los datos numéricos, es necesario convertir entre decimal y binario, almacenar números sin y con signo, y realizar operaciones aritméticas con ellos. Se presentan métodos de manipulación de números que escapan a la capacidad normal de la computadora. También se analizan la detección y corrección de errores.

## 55.1 NÚMEROS BINARIOS, OCTALES Y HEXADECIMALES

Todos estos sistemas numéricos, así como el sistema decimal, son *posicionales*, o sea, que el valor o peso asociado con una cifra o dígito depende de la posición y del valor de la cifra. Por ejemplo, en el sistema decimal el 3 del número 643 tiene el valor 3, mientras que en el número 634 tiene el valor 30. Primero se examinan números enteros, pero más adelante se extienden los conceptos para abarcar fracciones.

Un número decimal se puede desarrollar como una serie de potencias tal que el entero:

$$\dots d_3 d_2 d_1 d_0$$

puede expresarse como:

$$\dots + d_3 \times 10^3 + d_2 \times 10^2 + d_1 \times 10 + d_0$$

Así, el número 634 es:

$$6 \times 100 + 3 \times 10 + 4$$

En decimal, la base o raíz es 10 y cada una de las cifras de la serie  $d_i$  está en el intervalo  $0 \leq d_i < 10$ .

Un entero positivo cualquiera (mayor que 1) puede utilizarse como la raíz  $r$  de un sistema numérico, lo cual genera la expresión:

$$\dots + d_3 \times r^3 + d_2 \times r^2 + d_1 \times r + d_0$$

donde cada cifra  $d_i$  debe estar en el intervalo  $0 \leq d_i < r$ .

Las cuatro bases más utilizadas son:

Binario	$r = 2$	$0 \leq d_i \leq 1$
Octal	$r = 8$	$0 \leq d_i \leq 7$
Decimal	$r = 10$	$0 \leq d_i \leq 9$
Hexadecimal (hex)	$r = 16$	$0 \leq d_i \leq 15$

En todos estos sistemas se utilizan las cifras decimales estándares, salvo que en hexadecimal los números 10 a 15 se representan por medio de las letras A a F.

Para convertir de una base cualquiera a decimal se evalúa la serie, donde la aritmética se realiza en decimal.

## Ejemplos

$$\text{Binario } 1011 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2 + 1 = 8 + 2 + 1 = 11$$

$$10010 = 16 + 2 = 18$$

$$11111111 = 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1 = 255$$

$$111001011 = 256 + 128 + 64 + 8 + 2 + 1 = 459$$

$$\text{Octal } 471 = 4 \times 8^2 + 7 \times 8 + 1 = 313$$

$$77 = 7 \times 8 + 7 = 63$$

$$\text{Hexadecimal } 139 = 1 \times 16^2 + 3 \times 16 + 9 = 313$$

$$ABC = 10 \times 256 + 11 \times 16 + 12 = 2748$$

Para evitar dudas en cuanto a qué base se utiliza, a menudo ésta se escribe como subíndice, por ejemplo:

$$471_8 = 139_{16} = 313_{10}$$

En la tabla 55-1 se presentan las 32 primeras potencias positivas de 2 y  $1/2$  (que se necesitarán más adelante cuando se estudien las fracciones). Obsérvese que las potencias de 8 y de 16 también están incluidas en la tabla, ya que  $8^m = 2^{3m}$  y  $16^k = 2^{4k}$ .

## Ejemplo (de la tabla 55-1)

$$8^3 = 2^9 = 512$$

$$16^2 = 2^8 = 256$$

$$8^4 = 16^3 = 2^{12} = 4096$$

La tabla 55-2 incluye los 32 primeros enteros positivos en decimal, binario, octal y hexadecimal (así como en decimal codificado en binario, que se estudiará más adelante).

Los números decimales se analizan aquí porque son aquellos con los que todo el mundo está familiarizado. Se incluye el sistema binario debido a que el almacenamiento interno de todas las computadoras emplea dispositivos binarios. Aunque algunas máquinas codifican números decimales en binario, es esencial entender el sistema binario y las conversiones que se realizan entre los sistemas binario y decimal. Los sistemas octal y hexadecimal se incluyen porque son una notación abreviada del sistema binario, lo que permite escribir y recordar mejor los números binarios.

Para realizar la conversión entre binario y octal, sólo es necesario recordar que el octal representa la agrupación de cifras (dígitos) binarias, de tres en tres. Por

Tabla 55-1. Potencias de 2

$2^n$	$n$	$2^{-n}$
1	0	1.0
2	1	0.5
4	2	0.25
8	3	0.125
16	4	0.062 5
32	5	0.031 25
64	6	0.015 625
128	7	0.007 812 5
256	8	0.003 906 25
512	9	0.001 953 125
1 024	10	0.000 976 562 5
2 048	11	0.000 488 281 25
4 096	12	0.000 244 140 625
8 192	13	0.000 122 070 312 5
16 384	14	0.000 061 035 156 25
32 768	15	0.000 030 517 578 125
65 536	16	0.000 015 258 789 062 5
131 072	17	0.000 007 629 394 531 25
262 144	18	0.000 003 814 697 265 625
524 288	19	0.000 001 907 348 632 812 5
1 048 576	20	0.000 000 953 674 316 406 25
2 097 152	21	0.000 000 476 837 158 203 125
4 194 304	22	0.000 000 238 418 579 101 562 5
8 388 608	23	0.000 000 119 209 289 550 781 25
16 777 216	24	0.000 000 059 604 644 775 390 625
33 554 432	25	0.000 000 029 802 322 387 695 312 5
67 108 864	26	0.000 000 014 901 161 193 847 656 25
134 217 728	27	0.000 000 007 450 580 596 923 828 125
268 435 456	28	0.000 000 003 725 290 298 461 914 062 5
536 870 912	29	0.000 000 001 862 645 149 230 957 031 25
1 073 741 824	30	0.000 000 000 931 322 574 615 478 515 625
2 147 483 648	31	0.000 000 000 465 661 287 307 739 257 812 5

ejemplo, considérese el número binario 110101 (que corresponde al número decimal 53). Éste se expresa en forma de serie como:

$$1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2 + 1$$

o, agrupando términos de tres en tres y factorizando  $2^3=8$ , como:

$$8 \times (1 \times 4 + 1 \times 2 + 0) + (1 \times 4 + 0 \times 2 + 1)$$

o, por último:

$$8 \times 6 + 5$$

Cada término entre paréntesis es una cifra octal legítima (que puede asumir los valores desde 0, si las tres cifras binarias son 000, hasta 7, si las cifras binarias son 111). Si existen más bits\*, la serie tendrá más términos; el término que sigue es  $2^6$ , que se factorizará y producirá el término  $8^2$  de la serie octal. El método básico consiste en agrupar el número binario en grupos de tres bits, comenzando por el lado derecho (el bit menos significativo) y realizando la conversión de cada grupo a una cifra entre 0 y 7.

## Ejemplos (de binario a octal)

$$101110111 = 101\ 110\ 111 = 567$$

$$1101010 = 1\ 101\ 010 = 001\ 101\ 010 = 152$$

En la segunda serie, se agregaron ceros iniciales para completar el grupo de tres. (Esto, obviamente, no altera el valor del número.)

La conversión de octal a binario es simplemente una traducción dígito (cifra) por dígito. Cada cifra (dígito) octal se sustituye por su equivalente binario de tres bits.

## Ejemplos (de octal a binario)

$$713 = 111\ 001\ 011 = 111001011$$

$$177 = 001\ 111\ 111 = 1111111$$

Aquí, los espacios se presentan sólo para indicar el proceso, y en la segunda serie los ceros iniciales se eliminan.

Para convertir de binario a hexadecimal, los bits se agrupan de cuatro en cuatro, donde cada grupo se traduce a una cifra hexadecimal. En el proceso inverso, cada cifra hexadecimal se sustituye por cuatro bits.

## Ejemplos (de binario a hexadecimal)

$$10101001 = 1010\ 1001 = A9$$

$$101110111 = 0001\ 0111\ 0111 = 177$$

\* Bit es la contracción de las palabras *binary digit* (dígito binario).

$$2B0 = 0010\ 1011\ 0000 = 1010110000$$

Por último, para convertir de octal a hexadecimal (o viceversa), el procedimiento más sencillo consiste en convertir primero a binario, reagrupar los bits y después convertir a la otra base.

## Ejemplos (de octal a hexadecimal)

$$713_8 = 111\ 001\ 011 = 0001\ 1100\ 1011 = 1CB_{16}$$

$$3A0_{16} = 0011\ 1010\ 0000 = 001\ 110\ 100\ 000 = 1640_8$$

La conversión del sistema decimal a cualquiera de los otros sistemas de números también podría realizarse mediante la evaluación de una serie de potencias, pero la aritmética se deberá efectuar en el sistema numérico al cual se realiza la conversión (p. ej., binario u octal). Dado que pocas personas se sienten seguras al trabajar en bases diferentes a la decimal, se han desarrollado otros métodos que requieren solamente aritmética decimal. Para convertir a la base  $r$ , el número se divide entre  $r$ . El residuo es la cifra de orden inferior. La parte entera se vuelve a dividir entre  $r$ ; el residuo es la cifra siguiente. Este proceso continúa hasta que la parte entera se hace cero; en cada división se produce una cifra.

## Ejemplos (459<sub>10</sub>)

	Residuo		Residuo
459/2 = 229	1	459/8 = 57	3
229/2 = 114	1	57/8 = 7	1
114/2 = 57	0	7/8 = 0	7
57/2 = 28	1	459 <sub>10</sub> = 713 <sub>8</sub>	
28/2 = 14	0		
14/2 = 7	0		
7/2 = 3	1	459/16 = 28	11(B)
3/2 = 1	1	28/16 = 1	12(C)
1/2 = 0	1	1/16 = 0	1
459/10 = 111001011 <sub>2</sub>		459 <sub>10</sub> = 1CB <sub>16</sub>	

Obsérvese que estos resultados coinciden con las conversiones de algunos de los ejemplos anteriores. Se pueden utilizar dos atajos. Primero, el proceso de división puede finalizarse en un punto cualquiera cuando la parte entera se pueda convertir en forma directa. En la conversión de decimal a binario, por ejemplo, si se reconoce que  $14_{10}=1110_2$ , se puede suspender la división. Ésta produjo los bits 01011; el equivalente binario de 14, 1110, se agrega entonces al inicio. También es posible convertir de decimal a binario convirtiendo antes a octal y después de octal a binario. Esto requiere un tercio del número de divisiones y por tanto suele ser más rápido.

Antes de analizar las fracciones, se examinará brevemente la razón por la cual funciona este método. Considérese el número siguiente:

Tabla 55-2. Equivalentes decimales y binarios

Decimal	Binario	Octal	Hexadecimal	Decimal codificado en binario	
				Primera cifra	Segunda cifra
0	000000	00	00	0000	0000
1	000001	01	01	0000	0001
2	000010	02	02	0000	0010
3	000011	03	03	0000	0011
4	000100	04	04	0000	0100
5	000101	05	05	0000	0101
6	000110	06	06	0000	0110
7	000111	07	07	0000	0111
8	001000	10	08	0000	1000
9	001001	11	09	0000	1001
10	001010	12	0A	0001	0000
11	001011	13	0B	0001	0001
12	001100	14	0C	0001	0010
13	001101	15	0D	0001	0011
14	001110	16	0E	0001	0100
15	001111	17	0F	0001	0101
16	010000	20	10	0001	0110
17	010001	21	11	0001	0111
18	010010	22	12	0001	1000
19	010011	23	13	0001	1001
20	010100	24	14	0010	0000
21	010101	25	15	0010	0001
22	010110	26	16	0010	0010
23	010111	27	17	0010	0011
24	011000	30	18	0010	0100
25	011001	31	19	0010	0101
26	011010	32	1A	0010	0110
27	011011	33	1B	0010	0111
28	011100	34	1C	0010	1000
29	011101	35	1D	0010	1001
30	011110	36	1E	0011	0000
31	011111	37	1F	0011	0001

$$171_{10} = 253_8 = 2 \times 8^2 + 5 \times 8 + 3$$

Cuando se divide 171 entre 8, se obtiene 21 con un residuo de 3. En la serie de potencias de 8, cada uno de los términos, menos el primero, se dividen equitativamente; por tanto, el término de orden inferior deberá ser el residuo. El número entero resultante es:

$$\frac{2 \times 8^2}{8} + \frac{5 \times 8}{8} = 2 \times 8 + 5$$

Cuando se repite el proceso, el segundo término, 5, es el residuo; los otros términos se dividen en forma equitativa, lo que deja una parte entera de 2. La división final produce el número 2 como la tercera cifra octal.

La representación en serie de números se aplica a fracciones (valores no enteros) así como a enteros. Por tanto, el número decimal:

$$0.63 = 6 \times 10^{-1} + 3 \times 10^{-2}$$



El mismo desarrollo se aplica para otras bases, donde los términos son potencias negativas de la raíz  $r$  (o bien, en forma equivalente, potencias de  $1/r$ ). Para convertir un número a decimal, se evalúa la serie. (Puede utilizarse la tabla 55-1.)

**Ejemplos**

**Binario**  $.101 = 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} = .5 + .125 = .625$

$.0110011 = .25 + .125 + .015625 + .0078125 = .3984375$

**Octal**  $.5 = 5 \times 8^{-1} = .625$

$.23 = 2 \times 8^{-1} + 3 \times 8^{-2} = .296875$

**Hexadecimal**  $.A = 10 \times 16^{-1} = .625$

$.18 = 16^{-1} + 8 \times 16^{-2} = .09375$

Para convertir de decimal a una raíz cualquiera, se multiplica el número por la raíz. La parte entera del producto es la primera cifra de la respuesta. La parte fraccionaria se vuelve a multiplicar, lo que genera la segunda cifra. Este proceso continúa hasta que la fracción se hace cero.

**Ejemplos**

$.375 \times 2 = 0.75$   
 $.75 \times 2 = 1.50$   
 $.5 \times 2 = 1.00$   
 $.375_{10} = .011_2$   
 $.3828125 \times 8 = 3.0625$   
 $.0625 \times 8 = 0.50$   
 $.5 \times 8 = 4.0$   
 $.3828125_{10} = .304_8$   
 $.625 \times 16 = 10.0$   
 $.625_{10} = .A_{16}$

Para convertir de binario a octal (hexadecimal), se forman grupos de tres (cuatro) bits, comenzando en el punto de la raíz. De forma inversa, para convertir de octal (hexadecimal) a binario, se sustituye cada cifra por tres (cuatro) bits.

**Ejemplos**

$.011_2 = .3_8$   
 $.011_2 = .0110_2 = .6_{16}$   
 $.304_8 = .011\ 000\ 100_2$   
 $= .0110\ 0010_2 = .62_{16}$

En estos ejemplos, el proceso terminó debido a que la fracción llegó por último a 0. Esto no sucede siempre, como en el caso siguiente:

$.6 \times 2 = 1.2$   
 $.2 \times 2 = 0.4$   
 $.4 \times 2 = 0.8$   
 $.8 \times 2 = 1.6$   
 $.6 \times 2 = 1.2$

$.6_{10} = .1001100110011\dots$

Esto no debe ser una sorpresa, ya que la conversión de  $1/3$  a decimal es  $0.33333\dots$ , que es una fracción incommensurable. La repetición ocurre asimismo en octal y en hexadecimal; puede repetirse cada cifra o un grupo de cifras.

**Ejemplo**

$.6_{10} = .999\dots_{16} = .46314631\dots_8$

Cuando se van a convertir números mixtos, las partes entera y fraccionaria se convierten por separado.

**Ejemplos**

**A hexadecimal**  $25.625_{10}$   
 $25/16 = 1$  residuo 9  
 por tanto, entero = 19  
 $.625 \times 16 = 10.0$   
 por tanto, fracción = .A  
 $25.625_{10} = 19.A$

**A octal**  $212.630859375$   
 $212/8 = 26$  residuo 4  
 $26/8 = 3$  residuo 2  
 $3/8 = 0$  residuo 3  
 $.630859375 \times 8 = 5.046875000$   
 $.046875 \times 8 = 0.375$   
 $.375 \times 8 = 3.0$   
 $212.630859375_{10} = 324.503$

**Tabla 55-3. Sumador binario de un bit**

Acarreo del paso anterior	$a$	$b$	Acarreo al paso siguiente	Suma
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

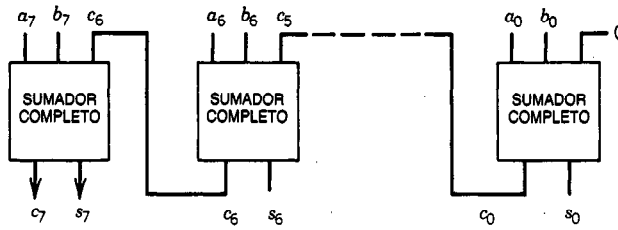


Fig. 55-1. Sumador binario de ocho bits.

## 55.2 SUMA BINARIA DE NÚMEROS POSITIVOS<sup>1</sup>

Los enteros binarios positivos se suman casi en la misma forma que los enteros decimales positivos. Se suma la cifra menos significativa de cada número, lo que produce una suma y un acarreo a la siguiente columna de cifras de orden superior. Después se suma el acarreo a esas cifras, lo que produce la segunda cifra de la suma y el acarreo a la tercera cifra. En la tabla 55-3 se presentan todas las posibilidades para una columna de la suma, donde la cifra del primer número se designa como *a* y la del segundo número como *b*.

### Ejemplo (de bit en bit)

	00	00	00	10
00011100	00011100	00011100	00011100	00011100
10111010	10111010	10111010	10111010	10111010
	0	1	1	0
11	11	01	00	
00011100	00011100	00011100	00011100	00011100
10111010	10111010	10111010	10111010	10111010
1	0	1	1	011010110

En este ejemplo se sumaron dos números de ocho bits y se produjo una suma de ocho bits. En realidad la suma es de nueve bits, pero el bit extra es cero y por tanto puede omitirse. Eso sucederá siempre que la suma sea 255 o menor (en este caso se sumaron 28 y 186 para generar una suma de 214). En cada paso del proceso, se suma el acarreo del paso anterior (que siempre es 0 la primera vez) al bit adecuado de cada número, lo que produce un bit de la suma más un acarreo al siguiente paso.

A continuación se presenta otro ejemplo.

### Ejemplo

1111111	
1011011	183
01001001	73
1 00000000	256

Obsérvese que la suma requiere nueve bits; es mayor de 255. (En realidad es posible llegar hasta 510 a

partir de enteros de ocho cifras, o 511 si hay un acarreo a la posición de orden inferior.) ¿Qué se hace con este bit extra? No hay espacio en la palabra para colocarlo.

Esta condición se conoce como desbordamiento. El **desbordamiento** se define como la generación de un resultado que escapa a la capacidad de la máquina. La mayor parte de las máquinas cuentan con un bit de acarreo para almacenar ese bit extra, de manera que el programador pueda hacerse cargo de la situación. (El desbordamiento no siempre es un error. Si se utilizan números de precisión múltiples, el desbordamiento es meramente un acarreo desde un grupo de ocho bits hacia el siguiente. Se volverá a abordar este tema en las secciones 55.4 y 55.6.) Enseguida se analiza un diagrama de hardware de un sumador de enteros positivos de ocho bits. En la figura 55-1 se presenta el diagrama de bloques de ese hardware. Cada bloque, rotulado como sumador completo, puede sumar un bit. Su comportamiento se describe en la tabla 55-3. El acarreo hacia el bit de orden inferior (bit 0) siempre es 0, y el acarreo desde el bit de orden superior es  $C_7$ , el bit extra de la suma.

## 55.3 NÚMEROS CON SIGNO

Cuando se utiliza un número de *k* cifras en raíz *r*, se pueden representar  $r^k$  valores. Por ejemplo, tres cifras decimales permiten  $10^3=1\,000$  representaciones, y ocho bits permiten  $2^8=256$  representaciones. Hasta este punto, estos números se han empleado para representar los enteros positivos del 0 al 999 en el caso de los números decimales de tres cifras y del 0 al 255 en el de los números binarios de ocho bits.

Cuando se representan números con signo, o sea, números que pueden ser positivos o negativos, aproximadamente la mitad de las representaciones disponibles se utilizan para valores positivos y la otra mitad para valores negativos. Esta distribución daría un intervalo de números de ocho bits de  $-127 \leq n \leq 127$  aproximadamente\*.

Se aplican dos métodos comunes para almacenar enteros con signo: el de magnitud y signo y el del comple-

\* El intervalo tendría aproximadamente 255 unidades, incluyendo el 0. Como hemos visto, en ocasiones la unidad 256 es desechada: otras veces el intervalo se extiende. Además, no es necesario mantener esta simetría. Se podría representar  $-55 \leq n \leq 200$  o cualquier otro intervalo de 256 números. Esta aproximación, puesto que raramente es utilizada, no será tratada aquí.

mento a dos. (También se ha utilizado el esquema del complemento a uno, pero no resulta muy popular.)

En el método de magnitud y signo, el bit inicial (principal) se reserva para el signo del número (0 en el caso de ser positivo y 1 si es negativo). Los bits restantes almacenan la magnitud. Los ejemplos que siguen ilustran este esquema de almacenamiento mediante el uso de ocho bits.

### Ejemplos

```

17 00010001
- 17 10010001
127 01111111
- 127 11111111
0 00000000
- 0 10000000
    
```

Como se puede apreciar en los ejemplos, el intervalo de números  $n$  que pueden representarse es  $-127 \leq n \leq 127$ . En general, una representación con magnitud y signo de  $k$  cifras tiene un intervalo de  $-(2^{k-1}-1) \leq n \leq (2^{k-1}-1)$ .

A pesar de las ventajas del método de magnitud y signo, dos desventajas han limitado su uso en las computadoras. Primero, existen dos representaciones del cero, una con un bit de signo positivo y otra con un bit de signo negativo. Esto complica el hardware que se utiliza para probar si un número es, por ejemplo, menor que cero, ya que ambos ceros se tratarán de igual forma. Un segundo problema ocurre en la suma de números en la forma de magnitud y signo; esto se analiza en la sección que sigue.

El complemento a dos es un sistema numérico modular diseñado para simplificar la aritmética. Por ahora el examen se limitará al formato de almacenamiento. Como sucede en el método de magnitud y signo, el bit inicial (principal) suele utilizarse como un bit de signo (0 si es positivo, 1 si es negativo). Los números positivos se almacenan en forma directa; por tanto, tienen exactamente el mismo formato que aquel método. En el caso de un sistema de ocho bits, se pueden almacenar enteros positivos hasta  $2^7-1=127$ .

Para la representación de un número negativo en un sistema numérico de complemento a dos se aplican varios métodos. Quizá el más simple (y el empleado en la mayor parte del hardware de computadoras) es el siguiente:

1. Se encuentra la representación de la magnitud del número.
2. Se sustituyen todos y cada uno de los bits por su complemento (comp).
3. Se suma 1 al resultado.

### Ejemplos

```

-1.    + 1  00000001
        Comp 11111110
        Sumar 1 11111111
        Almacenar 11111111
    
```

```

-17.    + 17  00010001
        Comp 11101110
        Sumar 1 11101111
        Almacenar 11101111
    
```

```

-0.      + 0  00000000
        Comp 11111111
        Sumar 1 11111111
        Almacenar 1 00000000
    
```

(ignorar el noveno bit acarreado; por tanto, sólo hay representación de 0)

```

-128.    + 128  10000000
        Comp 01111111
        Sumar 1 10000000
        Almacenar 10000000
    
```

Obsérvese que el número binario 10000000 podría representar +128 o -128. Suele elegirse el valor negativo, ya que el bit inicial (principal) es 1.

Una segunda técnica de conversión de números negativos consiste en sumar el número a  $2^k$  (256 para  $k=8$ ) y después convertir el resultado al sistema binario.

### Ejemplos

```

-1. 256 - 1 = 255
    255: 11111111
    
```

```

-17. 256 - 17 = 239
    239: 11101111
    
```

```

-0. 256 - 0 = 256
    256: 1 00000000
    (ignorar el noveno bit)
    
```

```

-128. 256 - 128 = 128
    128: 10000000
    
```

Como puede observarse en estos ejemplos, solamente existe una representación de 0 en complemento a dos, y el intervalo de valores negativos se extiende a -128, en comparación con el límite de -127 para el método de magnitud y signo. En general, una representación en complemento a dos de  $k$  cifras tiene un intervalo de  $-2^{k-1} \leq n \leq (2^{k-1}-1)$ .

La interpretación de los números sigue un patrón similar. Si la cifra inicial (principal) es 0, el número es positivo y es tan sólo la representación binaria directa. Si el bit inicial (principal) es 1, el número es negativo. Después se complementa cada bit y se suma 1. (Podría haberse restado 1 y después complementado cada bit para efectuar exactamente lo inverso del procedimiento de codificación; el resultado sería el mismo.)

### Ejemplos

```

01010110 + 86
10101010 Número negativo, complementar cada bit
    
```

01010101

$$\begin{array}{r} 01010101 \\ \underline{1} \end{array}$$

Sumar 1

01010110 Magnitud = 86; número = -86

10000000 Número negativo, complementar cada bit

01111111

$$\begin{array}{r} 01111111 \\ \underline{1} \end{array}$$

Sumar

10000000 Magnitud = 128; número = -128

En la tabla 55-4 se presenta una lista del valor de cada representación (mediante números de cinco bits) para enteros positivos y para enteros con signo (magnitud y signo y complemento a dos).

## 55.4 ARITMÉTICA BINARIA

Ahora se examinará la adición de números con signo. Ya se dijo que una de las desventajas del método de magnitud y signo es la dificultad de la adición. Para sumar dos números en representación de magnitud y signo, por ejemplo  $a$  y  $b$ , primero se debe determinar si coinciden sus signos. Si así sucede, se suman las magnitudes y se conserva el mismo signo. Si los signos son diferentes, se debe restar la menor magnitud de la mayor y conservar el signo de la mayor. En los ejemplos que siguen se ilustran algunos de los casos diferentes que podrían suceder en la adición.

**Tabla 55-4. Representación de números de cinco bits como enteros positivos, con magnitud y signo y complemento a dos**

Representación	Enteros positivos	Magnitud y signo	Complemento a dos
00000	0	+0	+0
00001	1	+1	+1
00010	2	+2	+2
00011	3	+3	+3
00100	4	+4	+4
00101	5	+5	+5
00110	6	+6	+6
00111	7	+7	+7
01000	8	+8	+8
01001	9	+9	+9
01010	10	+10	+10
01011	11	+11	+11
01100	12	+12	+12
01101	13	+13	+13
01110	14	+14	+14
01111	15	+15	+15
10000	16	-0	-16
10001	17	-1	-15
10010	18	-2	-14
10011	19	-3	-13
10100	20	-4	-12
10101	21	-5	-11
10110	22	-6	-10
10111	23	-7	-9
11000	24	-8	-8
11001	25	-9	-7
11010	26	-10	-6
11011	27	-11	-5
11100	28	-12	-4
11101	29	-13	-3
11110	30	-14	-2
11111	31	-15	-1

**Ejemplos**

+23 *a*  
+46 *b* Sumar las magnitudes, conservar el signo  
+69

-19 *a*  
-16 *b* Sumar las magnitudes, conservar el signo  
-35

+35 *a* Calcular  $|a| - |b|$   
-19 *b* Utilizar el signo de *a*  
+16

-19 *a* Calcular  $|b| - |a|$   
+35 *b* Utilizar el signo de *b*  
+16

-35 *a* Calcular  $|a| - |b|$   
+19 *b* Utilizar el signo de *a*  
-16

Para implementar esto se requiere un sumador, un sustractor y un comparador de la magnitud (para determinar qué número se va a restar de cuál).

Para sumar números en la representación de complemento a dos solamente se necesita sumar sus representaciones, ignorando el acarreo desde el último paso. Por tanto, es posible utilizar el mismo sumador binario de ocho bits. Después se ignora el acarreo desde el bit más significativo.

**Ejemplos**

30 00011110  
21 00010101  
51 00110011

3 00000011  
-2 11111110  
+1 (1) 00000001

-3 11111101  
2 00000010  
-1 11111111

-3 11111101  
-2 11111110  
-5 (1) 11111011

Obsérvese que en cada uno de estos casos la respuesta es correcta; el acarreo desde el último paso no es una indicación de desbordamiento. El desbordamiento se produciría si la suma generara un resultado fuera del intervalo, o sea, mayor que 127 o menor que -128. Los ejemplos que siguen ilustran estos dos casos.

**Ejemplos**

60 00111100  
100 01100100  
-96 10100000

-60 11000100  
-100 10011100  
+96 (1) 01100000

En cada caso se sumaron dos números del mismo signo y se produjo un resultado del signo contrario. Ésta es la indicación de desbordamiento. (El desbordamiento también puede determinarse a partir del acarreo hacia y desde el último paso.) Muchas computadoras cuentan con hardware para detectar el desbordamiento a partir de la suma de complemento a dos, y lo almacenan en un bit de desbordamiento especial.

La resta se efectúa con mayor frecuencia calculando primero el valor negativo del sustraendo y después sumando. El valor negativo suele obtenerse complementando cada bit del sustraendo y después sumando 1. La suma de 1, por lo general, se realiza como parte de la suma de dos números, colocando un 1 en el bit de acarreo de entrada de orden inferior. Los dos pasos (complemento y suma) se realizan con una sola instrucción de sustracción.

**Ejemplo.** Para calcular  $3 - 2$  se hace lo siguiente:

00000011 00000011  
-00000010 11111101 (complemento bit por bit)  
(1) 00000001

En la figura 55-2 se muestra un circuito para realizar la suma y resta de complemento a dos, mediante el mismo sumador empleado para enteros positivos, donde  $\text{sub}=1$  si la operación es la resta y  $\text{sub}=0$  si es la suma. Para la resta, el sustraendo se deberá complementar antes de realizar la suma. Obsérvese que el acarreo que produce la suma es el complemento de lo que se "pide prestado". En este ejemplo, lo que se pide prestado es 0, ya que 2 es menor que 3 y el bit de acarreo es 1. Algunas computadoras almacenan este bit en el bit de acarreo en una instrucción de resta; otras lo complementan y almacenan ahí lo que se pide prestado.

La multiplicación<sup>2</sup> se realiza mediante sumas sucesivas y desplazamiento. El producto siempre es de doble longitud; es decir, el producto de dos números de ocho bits es uno de 16. El proceso de multiplicación en binario es igual que en decimal. El multiplicando se multiplica sucesivamente por cada bit del multiplicador, se colocan escalonadamente los resultados, y se suman éstos. Desde luego, la multiplicación por un bit es simple, ya que multiplicar por 1 no altera el multiplicando, y multiplicar por cero produce sólo ceros.

**Ejemplos**

1101 (13)  
1011 (11)  
1101  
1101  
0000  
1101  
10001111 (143)

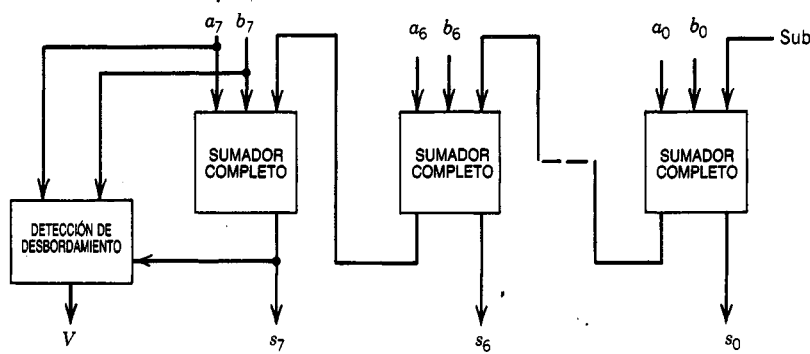


Fig. 55-2. Sumador-sustractor de complemento a dos.

La adición implica la suma de cuatro números en este ejemplo. Normalmente se realiza sumando cada renglón en un producto parcial conforme éste se calcula. Por tanto, una versión más descriptiva es:

$$\begin{array}{r}
 1101 \\
 \times 1011 \\
 \hline
 1101 \\
 11010 \\
 00000 \\
 0100111 \\
 \hline
 1101 \\
 10001111
 \end{array}$$

En cada adición sólo se suman números de cuatro bits. (Para el algoritmo de la computadora véase el cap. 59.)

La división<sup>2</sup> se efectúa por un proceso de restas de ensayo repetidas. Considérese el ejemplo de la figura 55-3, en el cual el producto del ejemplo de multiplicación (10001111) se divide entre 1011. Esta división termina después de cuatro pasos en el caso de estos números de cuatro bits. El residuo es 0000 y, por tanto, el cociente es un número entero.

### 55.5 DECIMAL CODIFICADO EN BINARIO (BCD)<sup>3</sup>

Hasta aquí se ha supuesto que cualquier número por almacenar en la computadora se introduce en binario (o alguna forma abreviada de binario) o se convierte de decimal a binario antes de que se almacene o se utilice. La salida se presenta en binario, o bien se re-convierte a decimal.

Dividendo: 10001111 Divisor: 1011

**Paso 1.** Restar el divisor de los cinco bits de la izquierda del dividendo. (La sustracción se puede realizar por resta real o complementando la división y sumando después):

$$\begin{array}{r}
 10001 \\
 1011 \\
 \hline
 0110
 \end{array}$$

Resta válida (resultado no negativo); el bit de orden superior del cociente es 1.

**Paso 2.** Agregar el siguiente bit del dividendo y volver a restar:

$$\begin{array}{r}
 01101 \\
 1011 \\
 \hline
 0010
 \end{array}$$

Resta válida; el siguiente bit es 1.

**Paso 3.** Agregar el siguiente bit y restar:

$$\begin{array}{r}
 00101 \\
 1011 \\
 \hline
 *
 \end{array}$$

Resultado negativo; el siguiente bit es 0. (\*La forma de la respuesta dependerá del hardware que se utilice para hacer la resta. Pero el resultado contendrá un indicador de número negativo.)

**Paso 4.** Restituir el dividendo al valor anterior (00101); agregar el último bit del dividendo; hacer la resta:

$$\begin{array}{r}
 001011 \\
 1011 \\
 \hline
 0000
 \end{array}$$

Resta válida; el último bit es 1:

**Respuesta:** 1011

Fig. 55-3. Ejemplo de división.

Tabla 55-5. Decimal codificado en binario

Cifra	Código binario
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

Para el ser humano no es natural trabajar en el sistema binario, aunque muchos sistemas pequeños (en particular los microprocesadores) están diseñados para operar en ese modo. La conversión requiere tiempo (un programa no trivial) o una cantidad de hardware considerable.

Otro método que se aplica en algunas máquinas consiste en codificar el número decimal en binario, de cifra en cifra. El procedimiento que se emplea más comúnmente consiste en sustituir cada cifra decimal por un número binario de cuatro bits (el equivalente binario), como se muestra en la tabla 55-5. (En ocasiones se emplean otros códigos, pero no se estudiarán aquí.)

#### Ejemplos

Decimal	BCD
74	0111 0100
29	0010 1001

El BCD tiene dos ventajas sobre el binario. Primero, como ya se ha visto, el proceso de conversión es mucho más simple. Esto es de particular importancia en aplicaciones tales como las calculadoras, donde constantemente se visualizan resultados intermedios en decimal. Segundo, elimina algunos errores introducidos en el proceso de conversión. Las fracciones que se pueden representar exactamente en decimal se pueden representar también con exactitud en BCD, pero no necesariamente en binario.

#### Ejemplo

**Decimal** .24

**BCD** .00100100

**Binario** .001111010111 (12 lugares)

**Decimal exacto del binario de 12 bits** .239990234375

Si el número binario de 12 bits se reconvierte a decimal (como en la última línea) en este ejemplo, el

resultado es incorrecto por 1 en el quinto lugar decimal. Esto puede causar problemas, en particular en cálculos financieros.

El BCD tiene dos desventajas importantes, y en ambas interviene el costo.

1. Requiere más espacio de almacenamiento. Para representar un número decimal de tres cifras se requieren 12 bits, cuatro para cada cifra. Este número de tres cifras proporciona 1 000 valores. Pero si se utilizara el sistema binario, se podrían representar 1 024 valores con sólo 10 bits. Por tanto, se requiere el 20% más de almacenamiento (12 bits en comparación con 10) para un intervalo de valores ligeramente menor.
2. La aritmética que normalmente se utiliza es más compleja y requiere hardware extra.

Antes de pasar a considerar la adición de números BCD, se analizará el almacenamiento de números con signo.

Como sucede en el caso del binario, en las computadoras se utiliza por lo general un sistema numérico de complemento. El equivalente del complemento a dos es el complemento a diez (de donde el término genérico es el de complemento a la raíz). Con  $k$  cifras ( $4k$  bits), se pueden realizar  $10^k$  representaciones. Dado que se emplea aproximadamente la mitad de esas para números positivos, el intervalo es  $-5 \times 10^{k-1} \leq n \leq (5 \times 10^{k-1} - 1)$ . Para  $k=2$ , se tiene  $-50 \leq n \leq 49$ , y para  $k=4$ ,  $-5\,000 \leq n \leq 4\,999$ .

Los números positivos se almacenan en forma directa. Como en binario, se dispone de dos métodos de cálculo para determinar la forma equivalente para almacenamiento de números negativos. En el primer método dicha forma se determina encontrando  $10^k + n$ . En el segundo, cada cifra se complementa con respecto a 9 (es decir, se resta de 9) y luego se suma 1 al número.

#### Ejemplos (cuatro cifras)

Número	Almacenamiento digital	Almacenamiento binario (16 bits)
5	0 005	0000 0000 0000 0101
1 234	1 234	0001 0010 0011 0100
-1 234	8 766	1000 0111 0110 0110
-5 000	5 000	0101 0000 0000 0000

En este método, la primera cifra de todos los números positivos (y 0) está entre 0 y 4, y la primera cifra de todos los números negativos está en el intervalo de 5 a 9.

Algunas veces resulta indeseable limitar el intervalo de la primera cifra. En este caso se agrega un bit extra como bit de signo. (Ahora se tienen  $4k+1$  bits, es decir, 17 bits. A menudo se hace referencia a esta cantidad como 4 1/2 cifras.) Por tanto, el intervalo de números se extiende hasta  $-10\,000 \leq n \leq 9\,999$ . En la primera operación, el número al cual se suma  $n$  se

convierte en 20 000. La segunda operación se mantiene sin cambio, salvo que la cifra del signo (sólo un bit) se complementa con respecto a 1.

### Ejemplos

Número	Almacenamiento digital	Almacenamiento binario
5	00005	0 0000 0000 0000 0101
-1234	18766	1 1000 0111 0110 0110
-9321	10679	1 0000 0110 0111 1001
+9999	09999	0 1001 1001 1001 1001

Ahora se analizará el proceso de adición de números BCD<sup>4</sup>. Dicho proceso suele ser de dos pasos. Primero se realiza la adición binaria y después, cuando es necesario, se suma un factor de corrección. Cuando se suman los códigos para una cifra del sumando y el

**Tabla 55-6. Sumador decimal codificado en binario (BCD)**

Sumador binario	Número representado	Representación correcta
C <sub>0</sub> * S <sub>1</sub> * S <sub>2</sub> * S <sub>3</sub> * S <sub>4</sub> *		C <sub>0</sub> S <sub>1</sub> S <sub>2</sub> S <sub>3</sub> S <sub>4</sub>
0 0 0 0 0	0 0	0 0 0 0 0
0 0 0 0 1	0 1	0 0 0 0 1
0 0 0 1 0	0 2	0 0 0 1 0
0 0 0 1 1	0 3	0 0 0 1 1
0 0 1 0 0	0 4	0 0 1 0 0
0 0 1 0 1	0 5	0 0 1 0 1
0 0 1 1 0	0 6	0 0 1 1 0
0 0 1 1 1	0 7	0 0 1 1 1
0 1 0 0 0	0 8	0 1 0 0 0
0 1 0 0 1	0 9	0 1 0 0 1
0 1 0 1 0	1 0	1 0 0 0 0
0 1 0 1 1	1 1	1 0 0 0 1
0 1 1 0 0	1 2	1 0 0 1 0
0 1 1 0 1	1 3	1 0 0 1 1
0 1 1 1 0	1 4	1 0 1 0 0
0 1 1 1 1	1 5	1 0 1 0 1
1 0 0 0 0	1 6	1 0 1 1 0
1 0 0 0 1	1 7	1 0 1 1 1
1 0 0 1 0	1 8	1 1 0 0 0
1 0 0 1 1	1 9	1 1 0 0 1
1 0 1 0 0	—	—
1 0 1 0 1	—	—
1 0 1 1 0	—	—
1 0 1 1 1	—	—
1 1 0 0 0	—	—
1 1 0 0 1	—	—
1 1 0 1 0	—	—
1 1 0 1 1	—	—
1 1 1 0 0	—	—
1 1 1 0 1	—	—
1 1 1 1 0	—	—
1 1 1 1 1	—	—

cosumando y después a ello se suma el acarreo de la cifra inmediata inferior, el resultado está en el intervalo de 00000 a 10011 (de 0 a 19). Como puede apreciarse en la tabla 55-6, los resultados son correctos cuando la suma es menor que 10, pero se necesita una corrección cuando la suma es mayor o igual a 10. Una inspección del resultado muestra que si se suma 6 (binario 00110) a la salida del sumador binario en estos casos, la respuesta es correcta.

### Ejemplos

5	0101	6	0110	9	1001
3	0011	4	0100	8	1000
8	1000	10	1010	17	1 0001
sin corrección		+6	110	+6	110
			1 0000		1 0111

En algunas máquinas la corrección la hace el hardware. Se utiliza una instrucción especial ADD DECIMAL (suma decimal), que invoca los circuitos de adición y corrección.

En muchos microprocesadores de ocho bits pueden almacenarse dos cifras decimales en una palabra. La suma decimal se realiza haciendo primero la suma binaria (con la instrucción ADD). El acarreo desde la cifra de orden inferior se guarda en el bit de medio acarreo H, y el acarreo desde la cifra de orden superior se guarda en C. La instrucción DECIMAL ADJUST (DAA, ajuste decimal) realiza entonces la corrección (recién mostrada) de cada cifra. Para la cifra de la derecha, H se considera el primer bit de la suma. Para la cifra de la izquierda, C es dicho bit.

### Ejemplo

Decimal	Almacenamiento	Suma binaria
29	29	0010 1001
-03	97	1001 0111
26	(1) 26	1100 0000
C = 0 H = 1		

**Corrección 1.** Cifra de la derecha (suma 10 000=16).

0000  
0110  
0110

**Corrección 2.** Cifra de la izquierda (suma 01100=12).

1100  
0110  
C = 1 0010

**Respuesta.** 0010 0110 (26).

Ambas correcciones las efectúa (en el orden en que se presentan) el hardware como resultado de una instrucción DAA. Si la primera corrección produce un acarreo, éste se suma a la segunda cifra antes de realizarse la corrección.



**Ejemplo**

Decimal	Almacenamiento	Suma binaria
-23	77	0111 0111
<u>26</u>	<u>26</u>	<u>0010 0110</u>
3	(1) 03	1001 1101 C=0 H=0
Cifra de la derecha		<u>0110</u>
(01101 = 13), correcto		1010 0011
Cifra de la izquierda		
(01010 = 10), correcto		<u>0110</u>
Decimal: 03	(1) 000 0011	

5 6  
CD  
1 2 3

y después se suma con el acarreo:

1  
3 4  
AB  
0 E 0

Este mismo proceso puede aplicarse a números que requieran tres o más palabras de precisión. Las palabras de orden inferior se suman mediante el uso de ADD; después se suma cada una de las palabras de orden superior, a su vez, mediante el uso de ADC. La precisión múltiple se puede aplicar asimismo a la adición y la sustracción decimales.

### 55.6 ARITMÉTICA DE PRECISIÓN MÚLTIPLE

El tamaño estándar de los datos para operaciones aritméticas es la palabra. El número de bits en una palabra varía desde 4 en algunos microprocesadores hasta 64 en algunas computadoras de mayor tamaño. Salvo en el caso de los tamaños de palabra más grandes, es necesario manipular números de mayor precisión de la que se dispone en una sola palabra. Para ello se aplican dos métodos básicos. En algunas máquinas, se proporcionan instrucciones para aritmética de diversos tamaños de datos. Por ejemplo, en la Motorola 68000 se aceptan tres tipos de datos: el byte (ocho bits), la palabra (16 bits) y la palabra larga (32 bits). Las instrucciones aritméticas pueden operar con los tres tipos.

En muchas máquinas<sup>5</sup> (ya se utilice el primer método o no) se proporcionan instrucciones para ampliar la precisión de la suma y la resta a través del software. Para facilitar la exposición, aquí se supondrá un tamaño de palabra de ocho bits, aunque el concepto se aplica igualmente bien a otros tamaños de palabra. Cuando se calcula la suma de dos números de ocho bits, se obtiene un resultado de nueve bits, donde el noveno bit suele almacenarse como un bit de acarreo. Si los números fuesen más largos, el noveno bit sería tan sólo el acarreo hacia el siguiente bit superior. Para sumar dos números de 16 bits (en una máquina de ocho bits), primero se deben sumar los ocho bits inferiores de cada número. La suma se guarda como los ocho bits inferiores de la respuesta. Después se suman los ocho bits siguientes de las dos palabras, junto con el acarreo que produce la primera adición. Esa suma se guarda como los ocho bits superiores de la respuesta. Se incluye una instrucción especial, ADD WITH CARRY (ADC, suma con acarreo), que suma el bit de acarreo a los dos números.

**Ejemplo (en hexadecimal).** Para sumar:

3 4 5 6  
ABCD  
E 0 2 3

primero se suma:

### 55.7 ARITMÉTICA DE PUNTO FLOTANTE<sup>6,7</sup>

Podría diseñarse una computadora con números de punto fijo exclusivamente, y forzar a todo el mundo a escalar sus cálculos de manera que solamente trabajarán con enteros o sólo con fracciones; o hasta podría colocarse el punto binario siempre en el centro. De este modo, si se tuvieran palabras de 32 bits, podría utilizarse un bit para el signo, 16 a la izquierda del punto binario y 15 a la derecha de éste. Pero esto sólo permite números hasta  $2^{16}$  o aproximadamente 65 000, y apenas cuatro espacios decimales a la derecha del punto decimal. Por tanto, no sería posible representar (más que como cero) el número 0.00001, ni casi un millón. En cálculos científicos no son raros números tan grandes como  $10^{25}$ , y tampoco lo son números tan pequeños como  $10^{-20}$ . Para representar ambas cantidades con números de punto fijo se necesitarían 150 bits, un número mucho mayor que el disponible en una computadora cualquiera. En consecuencia, se ha introducido la notación de punto flotante en casi todas las computadoras científicas para hacer posible la representación de un mayor intervalo de números en una palabra de 32 bits.

Enseguida se presenta el punto flotante en decimal y después se explica cómo se almacenan los números en binario. Se hace esto sólo porque los números decimales son mejor conocidos. Con cinco cifras decimales más dos bits de signo, ahora es posible representar números en el intervalo de magnitudes  $.001 \times 10^{-99} \leq X \leq .999 \times 10^{99}$ . Aquí se utilizan tres cifras más un bit de signo para la *mantisa*, las cifras significativas, y dos cifras y un bit de signo para el *exponente*, la potencia de 10. En este formato (que es común), la mantisa es una fracción. Esta forma proporciona un intervalo más que suficiente para realizar casi cualquier cálculo, mientras que un número de punto fijo de cinco cifras no podría siquiera representar un millón; además, este esquema de representación en particular es inadecuado para muchos fines debido a la falta de exactitud. Tres espacios no suelen bastar. Por otra parte, una palabra de 32 bits suele asignar ocho o nueve bits al

exponente (incluyendo su signo) y los 23 o 24 bits restantes a la mantisa. Esto da aproximadamente siete cifras representativas, que son adecuadas para muchos cálculos. (Algunas máquinas pueden realizar operaciones de punto flotante de doble precisión en casos en que ésta es necesaria.)

Para conservar la precisión, los números de punto flotante casi siempre se almacenan con la primera cifra de la mantisa como un número distinto de cero (salvo cuando el número es exactamente cero). Así, por ejemplo, 23 se almacenaría como  $.230 \times 10^2$  en vez de como  $.023 \times 10^3$ . Esto reduce el intervalo ligeramente (hasta  $.100 \times 10^{-99}$ ), pero ello no implica mayor problema.

Ahora se analizará la forma en que podrían sumarse dos números de punto flotante, comenzando con un ejemplo sencillo:

$$\begin{array}{r} -.231 \times 10^1 \\ \underline{.642 \times 10^1} \\ .411 \times 10^1 \end{array} \quad \begin{array}{r} -2.31 \\ \underline{6.42} \\ 4.11 \end{array}$$

En este caso simplemente se sumaron las mantisas; no fue necesario hacer nada más. La primera complicación aparece cuando los dos números que se van a sumar tienen exponentes distintos. Por ejemplo:

$$\begin{array}{r} .412 \times 10^1 \\ \underline{.264 \times 10^2} \\ 30.52 \end{array}$$

No es posible sumar simplemente las mantisas, puesto que los exponentes son distintos; se deberá desplazar una de aquéllas. En particular, es necesario ajustar los números desplazando la mantisa del que tenga el menor exponente a la derecha tantos lugares como difieran los exponentes. Después se incrementa el exponente de ese número en el número de lugares recorridos. Esto deja el valor sin cambio, salvo en la operación de redondeo, es decir,  $.410 \times 10^1 = .041 \times 10^2$ . Aquí, como los exponentes difieren en 1, la primera mantisa se debe desplazar un lugar a la derecha, como sigue:

$$\begin{array}{r} .41 \times 10^2 \\ \underline{.264 \times 10^2} \\ .305 \times 10^2 \end{array}$$

Sólo es posible sumar dos mantisas cuando se han desplazado hasta que los exponentes son iguales. Existe la posibilidad de que el número menor sea tan pequeño que desaparezca cuando se desplace. Considérese por ejemplo:

$$\begin{array}{r} .237 \times 10^{21} \\ \underline{+.641 \times 10^{16}} \end{array}$$

Es necesario desplazar el segundo número cinco lugares a la derecha; eso produce  $.000 \times 10^{21}$ , y la suma es el primer operando.

Ahora se analizarán otras dos complicaciones. Primero, considérese:

$$\begin{array}{r} -.581 \times 10^2 \\ \underline{-.723 \times 10^2} \\ -1.304 \times 10^2 \end{array}$$

La suma que se produce es demasiado grande; requiere una cifra a la izquierda del punto decimal. Sin embargo, *no* es un desbordamiento. Es posible ajustar el resultado desplazando el número un lugar a la derecha (o el punto un espacio a la izquierda) y sumando 1 al exponente; es decir, la suma es:

$$-.130 \times 10^3$$

Por tanto, algunas veces es necesario ajustar la suma desplazándola un lugar a la derecha. El único caso en que se obtendrá un desbordamiento es cuando el exponente ya estaba en su valor máximo antes del ajuste, o sea, cuando el resultado es mayor que  $10^{99}$ . Entonces se tiene en verdad un desbordamiento.

El otro ajuste que se podría requerir es cuando se suman dos números que son casi iguales pero de signo contrario. Por ejemplo, considérese:

$$\begin{array}{r} .653 \times 10^2 \\ \underline{-.621 \times 10^2} \\ .032 \times 10^2 \end{array}$$

Pero esta suma no está en la forma correcta: la primera cifra es cero. En consecuencia, debe normalizarse el número desplazándolo un lugar a la izquierda y reduciendo el exponente en uno. Por tanto la forma correcta es:

$$.320 \times 10^1$$

Podría requerirse un ajuste de más de un espacio. Por ejemplo:

$$\begin{array}{r} -.511 \times 10^{-1} \\ \underline{.509 \times 10^{-1}} \\ -.002 \times 10^{-1} \\ \underline{.200 \times 10^{-3}} \end{array}$$

Es oportuno hacer aquí una advertencia. Si la suma es cero, podrían hacerse desplazamientos en forma indefinida sin obtener una primera cifra diferente de cero. Debe verificarse esto. Otro problema ocurre cuando debe hacerse esta corrección y el exponente se encuentra ya en su límite inferior (en  $-99$ ). En este caso se obtiene un número más pequeño del que puede representarse; esto se llama *insuficiencia*. Esta condición de error puede evitarse en la misma forma que el desbordamiento, o con frecuencia resulta más significativo simplemente hacer el resultado igual a 0. Después de todo, la magnitud está más próxima a cero de lo que es posible representar.

Ahora se analizará la forma en que se almacenan los números de punto flotante en una computadora. Todo lo que se necesita es la mantisa (con su signo) y el exponente (y su signo). El número se almacena como una fracción binaria (la mantisa) con la potencia de 2 (el exponente) adecuada o como una fracción hexade-

S	EXPONENTE	MANTISA
---	-----------	---------

Fig. 55-4. Formato de punto flotante.

cimal con la potencia de 16 adecuada. Por ejemplo, considérese el almacenamiento de 23.75. Primero se hace la conversión a binario:

$$23.75 = 10111.11 \\ = .1011111 \times 2^5$$

Por tanto, en un esquema binario, la mantisa sería +.1011111 y el exponente +101 (5 binario).

En hexadecimal se tendría:

$$23.75 = 17.C_{16} \\ = .17C \times 16^2$$

Como los números hexadecimales se almacenan en realidad como binarios, se tiene un almacenamiento de:

$$+00101111100 \quad y \quad +0010$$

¿Cuál es la diferencia entre los dos esquemas? Obsérvese que lo que finalmente se almacena como la mantisa es exactamente lo mismo en ambos casos, salvo por los ceros iniciales (o principales). En binario nunca hay un 0 inicial; en hexadecimal la inicial *cifra* (principal) no puede ser 0; es decir, nunca puede haber cuatro ceros binarios iniciales. Lo que se ahorra al utilizar el sistema hexadecimal es el número de bits que se requieren para representar el exponente. Para representar números hasta cerca de  $2^{256}$ , se necesitan ocho bits del exponente (más el signo). Por tanto, es posible ahorrar un poco en el tamaño del exponente cambiando al sistema hexadecimal. Pero al mismo tiempo se pierde un poco de exactitud debido a los ceros iniciales (no significativos). (El número promedio de ceros iniciales en hexadecimal, suponiendo que todas las cifras tienen la misma probabilidad, es .73. Por tanto, se tiene una ganancia de poco más de un bit de precisión.)

La mantisa casi siempre se almacena en el mismo sistema numérico que los enteros, ya que se requieren los mismos tipos de operaciones (a menudo mediante el uso del mismo hardware). El exponente puede tener un formato diferente, o bien se puede almacenar en el mismo sistema que la mantisa. Las razones de considerar un esquema diferente para el exponente son que se requiere una serie de operaciones más limitada y que el número 0 se puede representar en forma más efectiva. Un esquema común de almacenamiento del exponente en, por ejemplo, un formato de nueve bits, incluyendo el signo, consiste en almacenar el exponente,  $e$ , como:

$$2^8 + e$$

donde  $e$  deberá estar en el intervalo  $-256 = -2^8 \leq e \leq 2^8 - 1 = 255$ .

Éste se denomina algunas veces formato de exceso de  $256(2^8)$ . Así, se tiene por ejemplo:

	Almacenamiento
0	10000000
1	10000001
-1	01111111
255	11111111
-256	00000000

Obsérvese que todos los exponentes positivos (incluso 0) comienzan con 1 y todos los exponentes negativos, con un 0. Además, si los nueve bits se tratan como un entero binario positivo, cuanto mayor es ese entero mayor es el exponente. Asimismo, si se hace una resta mediante aritmética binaria normal (el menor se sustrae del mayor), se obtiene la respuesta correcta. La otra razón importante para utilizar este esquema es que todos los ceros de la mantisa y del exponente representan un cero de punto flotante verdadero. Esto es deseable, puesto que hace que el cero flotante y el entero cero sean el mismo y simplifica considerablemente las pruebas de 0. Por otra parte, si se utilizara un esquema de complemento, se tendría la opción (en complemento a dos):

$$.000 \times 2^{-256}$$

que tendría un exponente de 10000000 (el cual difiere de la representación entera de cero) o un exponente todo ceros, que sería:

$$.000 \times 2^0$$

Eso tampoco es muy satisfactorio, ya que el siguiente número en magnitud que se puede representar,  $.100 \times 2^{-256}$ , tiene un formato de almacenamiento muy diferente.

Un formato común para los números de punto flotante se muestra en la figura 55-4, donde S es el bit del signo de la mantisa. El signo, incluyendo el exponente, suele tener ocho o nueve bits, y el resto es para la mantisa.

**Ejemplo (IBM 370).** En el sistema IBM 370, el número de punto flotante tiene el formato de la figura 55-4, donde el exponente es  $64(2^6 + e = 64 + e)$  en exceso de siete bits, la mantisa es una fracción de 24 bits, y el bit S es el signo de la mantisa. El valor decimal del número está dado por:

$$\pm \text{mantisa} \times 16^{(e-64)}$$

El número +1 se almacena como  $16^{-1} \times 16^{65}$  o, en hexadecimal (mantisa=.10000 y  $e=65=41_{16}$ ):

$$41100000$$

La forma de 32 bits tiene valores que varían de:

$$16^{-65} \text{ a } (1 - 16^{-6}) \times 16^{63}$$

### 55.8 DETECCIÓN Y CORRECCIÓN DE ERRORES

Si el medio de almacenamiento o transmisión está propenso a errores, suele ser necesario agregar bits a la información para hacer posible la detección y corrección de los errores<sup>2,6,8</sup>.

La palabra se puede extender en un solo bit, a menudo llamado bit de paridad, de manera que se pueda detectar cualquier error que haya en esa palabra (incluso el bit de paridad). El bit de paridad se elige de manera que el número total de unos en la palabra extendida sea par. Si se comete un solo error en el procesamiento de la información, es decir, si un 1 se convierte en 0 o un 0 se convierte en 1, entonces el número total de unos ahora es impar. Por tanto, existe un error si hay un número impar de unos en la palabra extendida. Este método detectará un error en una palabra (o cualquier número impar de errores, 3, 5, etc.) pero no detectará un número par de errores. No obstante, mejora considerablemente la confiabilidad si la probabilidad de cometer un error es pequeña y los errores son independientes. Por ejemplo, si se suma un bit de paridad a una palabra de ocho bits y la probabilidad de que haya un error en un bit cualquiera es 0.1, entonces la probabilidad de que haya un error no detectado en una palabra es .00336. Sin detección de errores, la probabilidad de que haya un error es .0773 (23 veces mayor) y todos los errores pasan inadvertidos. Sin embargo, obsérvese que la adición de un bit de verificación reduce la probabilidad de que se realice una transmisión correcta (de .923 a .913), ya que ahora existen nueve bits en los cuales se puede cometer un error, en vez de ocho.

#### Ejemplos

Datos	Palabra extendida (cifra de verificación del lado derecho)
0 1 1 0 1 0 1 1	0 1 1 0 1 0 1 1 1
0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 0

Almacenados o transmitidos	Recuperados o recibidos	Error detectado
0 1 1 0 1 0 1 1 1	0 1 1 0 1 0 1 1 1	No
0 1 1 0 1 0 1 1 1	1 1 1 0 1 0 1 1 1	Sí
0 1 1 0 1 0 1 1 1	0 1 1 0 1 0 1 1 0	(Sí; datos correctos; error en el bit de paridad)
0 1 1 0 1 0 1 1 1	1 1 1 1 1 0 1 1 1	No (error doble no detectado)

En la presente descripción de la paridad se considera un número par de unos en la palabra extendida. El

**Tabla 55-7. Verificaciones de la paridad de ocho bits de datos con el código de Hamming**

Bit de verificación	Bit de datos							
	7	6	5	4	3	2	1	0
3	X	X		X	X		X	
2	X		X	X		X	X	
1		X	X	X				X
0					X	X	X	X

resultado no se modificaría en absoluto si el bit de paridad se eligiera de modo que el número total de bits fuera impar.

Cabe aquí hacer una advertencia con respecto a la terminología. Los términos "paridad impar" y "paridad par" se utilizan en forma inconsistente en la literatura. El término paridad par se utiliza algunas veces para indicar que el número total de unos en la palabra extendida es par (como en nuestro ejemplo). Pero otras veces se emplea para decir que el bit de paridad sería 1 si hay un número par de unos en los datos (que es exactamente lo contrario).

Un código de detección de errores solamente puede indicar que se ha cometido un error en alguna parte de la palabra; no indica qué bit tiene el error. Un código de corrección de errores corrige el error en forma automática.

En las tablas 55-7 y 55-8 se ilustra un código que corrige todos los errores individuales, llamado código de Hamming. Los cuatro bits de verificación se agregan a los bits de datos. Cada bit de verificación se calcula de modo que se verifique la paridad (par o impar) de los bits que se indican con una X en la tabla 55-7. Para detectar errores, se efectúan las cuatro verificaciones de paridad. Cada posible error individual

**Tabla 55-8. Diagrama de corrección con el código de Hamming**

Falla verificada (bit de verificación)				Bit en error
3	2	1	0	
	Ninguna			Ninguno
			X	C0
		X		C1
		X	X	D0
	X			C2
	X		X	D2
	X	X		D5
X				C3
X			X	D3
X		X		D6
X	X			D7
X	X		X	D1
X	X	X		D4

**Tabla 55-9. Bits necesarios de verificación para detectar errores individuales con una longitud de palabra dada**

Bits de datos	Bits de verificación
1	2
2 a 4	3
5 a 11	4
12 a 26	5
27 a 57	6
58 a 120	7

produce un patrón de fallas único. La tabla 55-8 contiene un diagrama de corrección, que relaciona las verificaciones falladas con el error.

**Ejemplos (número par de unos, incluyendo el bit de paridad)**

Datos: 0 0 1 0 1 1 0 1 Verificación: 1 0 0 1  
Palabra total: 0 0 1 0 1 1 0 1 1 0 0 1

Error 1: 0 1 1 0 1 1 0 1 1 0 0 1 (en el bit de datos 6)  
Verificaciones: falla, pase, falla, pase D6  
Datos corregidos: 0 0 1 0 1 1 0 1

Error 2: 0 0 1 0 1 1 0 1 0 0 0 1 (en el bit de verificación 3)  
Verificaciones: falla, pase, pase, pase C3  
Datos corregidos: 0 0 1 0 1 1 0 1

Error 3: 0 0 1 0 0 0 0 1 1 0 0 1 (en los bits de datos 3 y 2)  
Verificaciones: falla, falla, pase, pase D7  
Datos corregidos: 1 0 1 0 0 0 0 1 (ahora hay tres errores).

Obsérvese que en el último ejemplo se produce un doble error en un patrón que indica un error en un bit correcto. Como se utilizan 12 bits en vez de ocho, la probabilidad de que no haya error se reduce a .886. Pero todos los errores individuales se pueden corregir. Por tanto, la probabilidad de un error no corregido es solamente de .0062. Además, si se agrega un decimotercer bit (quinto bit de verificación) para proporcionar paridad global, se pueden detectar errores dobles, ya que todos éstos generarán una o más discrepancias en las primeras cuatro verificaciones, pero un quinto bit correcto. En tal caso, la probabilidad de un error sin corregir es .00725 y de un error inadvertido es .00027. Para ello, se utilizarán 13 bits en vez de 8.

Se pueden aplicar los mismos conceptos de detección y corrección de errores en el caso de palabras de

**Tabla 55-10. Código ASCII de siete bits**

				$b_6b_5b_4$							
$b_3$	$b_2$	$b_1$	$b_0$	000	001	010	011	100	101	110	111
0	0	0	0	NUL <sup>a</sup>	DLE	SP	0	@	P		p
0	0	0	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	STX	DC2	"	2	B	R	b	r
0	0	1	1	ETX	DC3	#	3	C	S	c	s
0	1	0	0	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	ACK	SYN	&	6	F	V	f	v
0	1	1	1	BEL	ETB	'	7	G	W	g	w
1	0	0	0	BS	CAN	(	8	H	X	h	x
1	0	0	1	HT	EM	)	9	I	Y	i	y
1	0	1	0	LF	SUB	*	:	J	Z	j	z
1	0	1	1	VT	ESC	+	;	K	[	k	{
1	1	0	0	FF	FS	,	<	L	\	l	
1	1	0	1	CR	GS	-	=	M	]	m	}
1	1	1	0	SO	RS	.	>	N	^	n	~
1	1	1	1	SI	US	/	?	O	—	o	DEL

<sup>a</sup> Caracteres de control: NUL, nulo; SOH, inicio de encabezado; STX, inicio de texto; ETX, fin de texto; EOT, fin de transmisión; ENQ, solicitud; ACK, reconocimiento; BEL, campana; BS, retroceso de un espacio; HT, tabulador horizontal; LF, avance de renglón; VT, tabulador vertical; FF, avance de forma; CR, regreso de carro; SO, mayúsculas; SI, minúsculas; DLE, cambio de significado (control); DC1, control 1 de dispositivo; DC2, control 2 de dispositivo; DC3, control 3 de dispositivo; DC4, control 4 de dispositivo; NAK, reconocimiento negativo; SYN, sincronización; ETB, fin del bloque de transmisión; CAN, cancelación; EM, fin del medio; SUB, sustituto; ESC, escape; FS, separador de archivos; GS, separador de grupos; RS, separador de registros; US, separador de unidades; SP, espacio; DEL, supresión (borrado).

mayor tamaño. En la tabla 55-9 se muestra el número de bits de verificación que se requieren a fin de detectar errores individuales para diversas longitudes de palabra. (Se necesita otro bit de verificación si también se requiere la detección de errores dobles.)

### 55.9 REPRESENTACIONES DE DATOS ALFANUMÉRICOS<sup>5,8,9</sup>

A menudo se introduce y se procesa en una computadora información textual. Ésta se deberá codificar en binario. Aunque se han utilizado muchos códigos, el más común actualmente es el código ASCII (*American Standard Code for Information Interchange*) de 7 bits, que se ilustra en la tabla 55-10. Además de letras mayúsculas y minúsculas, números y símbolos aritméticos y de puntuación especiales, cuenta con varios códigos para controlar la transmisión o la impresión de información. Como se indica en la tabla, la letra *H* se representaría como  $1001000=48_{16}$ .

Típicamente, cada carácter se almacena en un byte (ocho bits) de memoria. Por lo general se agrega un bit de paridad (sec. 55.8) al código de siete bits.

### REFERENCIAS BIBLIOGRÁFICAS

1. H. Taub, *Digital Circuits and Microprocessors*, McGraw-Hill, Nueva York.
2. R. K. Richards, *Arithmetic Operations in Digital Computers*, D. Van Nostrand, Nueva York.
3. V. Rajaraman y T. Radhakrishnan, *An Introduction to Digital Computer Design*, 2.<sup>a</sup> ed., Prentice-Hall, Englewood Cliffs, NJ.
4. R. L. Krutz, *Microprocessors and Logic Design*, Wiley, Nueva York.
5. S. B. Newell, *Introduction to Microcomputing*, Harper & Row, Nueva York.
6. T. C. Bartee, *Digital Computer Fundamentals*, 6.<sup>a</sup> ed., McGraw-Hill, Nueva York.
7. M. R. Sloan, *Computer Hardware and Organization*, 2.<sup>a</sup> ed., Science Research Associates, Chicago.
8. D. D. Givone y R. P. Roesser, *Microprocessors/ Microcomputers: An Introduction*, McGraw-Hill, Nueva York.
9. J. W. Gault y R. L. Pimmel, *Introduction to Microcomputer-Based Digital System*, McGraw-Hill, Nueva York.



## CAPÍTULO 56

# Organización y arquitectura de la computadora

**Edward J. Lancevich**  
Robotic Vision Systems, Inc.  
Hauppauge, Nueva York

- |   |  |
|---|--|
| <b>56.1 Organización general (diagrama de bloques)</b>            | <b>56.6 Diseño de la unidad de control</b>               |
| <b>56.2 Registros de la CPU</b>                                   | 56.6.1 Diseño de la CU por cableado                      |
| <b>56.3 Tipos de instrucciones</b>                                | 56.6.2 Diseño de la CU por microprogramación             |
| <b>56.4 Formatos de instrucciones y modos de direccionamiento</b> | <b>56.7 Clasificación de los sistemas de computación</b> |
| <b>56.5 Modos de operación</b>                                    | <b>56.8 Comparación de computadoras</b>                  |



## 56.1 ORGANIZACIÓN GENERAL (DIAGRAMA DE BLOQUES)

La organización funcional de un sistema de computación general se muestra en forma de diagrama de bloques en la figura 56-1. Los bloques funcionales básicos son la unidad de procesamiento central (CPU), la memoria principal y el procesador de entrada/salida (I/O).

**Unidad de procesamiento central.** La CPU es responsable de la interpretación y ejecución de instrucciones contenidas en la memoria principal. Las comunicaciones entre la CPU y la memoria principal se realizan a través de dos ductos funcionalmente distintos: el de direcciones y el de datos.

Para introducir en la memoria una instrucción específica, la CPU envía a aquella la dirección de la instrucción por el ducto de direcciones, y recibe por el mismo medio la instrucción que está en esa dirección. Parte de la instrucción es utilizada por la CPU para identificar la operación por realizar; esta parte se llama código de operación de la instrucción. La información restante se utiliza para determinar la localidad o las localidades de los datos con los cuales se va a efectuar la operación.

La acción de leer una instrucción en la CPU y prepararla para su ejecución se denomina ciclo de búsqueda (o de captura) de la computadora. Para completar una instrucción, la CPU decodifica el código de operación, genera las señales de control que se necesitan para introducir los operandos requeridos y controla la ejecución de la instrucción.

Por ejemplo, supóngase que la operación especificada consiste en sumar dos números retenidos en dos registros de la CPU y almacenar el resultado en un tercer registro de la CPU. Para ejecutar esta instrucción, la CPU identificaría los dos registros y generaría las señales de control adecuadas para conectar los registros a la unidad de aritmética y lógica (ALU) de la CPU. La CPU también haría que la ALU funcionara como sumador, y dirigiría la salida de la ALU hacia el tercer registro.

El proceso de realización de las operaciones que especifica una instrucción se denomina ciclo de ejecución<sup>1,2</sup> de la computadora.

Los nombres "ciclo de búsqueda" y "ciclo de ejecución" se derivan de la naturaleza cíclica de la operación de la computadora; una vez que ésta comienza a operar, repite los ciclos de búsqueda y ejecución de manera continua. Para hacer referencia a ambos ciclos en conjunto suele utilizarse el término "ciclo de máquina"<sup>1,2</sup>.

En el ejemplo está implícita la noción de que la CPU puede dividirse funcionalmente en tres subunidades: la unidad de control, dedicada a los ciclos de búsqueda y ejecución; la ALU, que desempeña funciones aritméticas (p. ej., suma y resta) y de lógica (p. ej., AND, OR); y un conjunto de registros, dedicados al almacenamiento de datos en la CPU y a ciertas funciones de control. Estas tres unidades se detallan en secciones posteriores.

**Memoria principal.** Consta de un conjunto de cadenas de bits\*. Una o más de estas cadenas de bits constituyen una instrucción de la computadora. La longitud (el número de bits) de una cadena se denomina longitud de palabra<sup>1,2</sup> de la computadora y es la más pequeña unidad independiente claramente direccionable contenida en la memoria de la máquina. El número máximo de cadenas que puede ser recibido por la computadora se conoce como espacio direccionable (o de direcciones) de la máquina.

Dado que las direcciones son números binarios, el espacio de direcciones es una potencia entera de dos; es decir,  $2^k$ , donde  $k$  representa el número de líneas en el ducto de direcciones.

Típicamente, el espacio de direcciones de una computadora se designa con la notación  $XK$  palabras; en esta notación,  $X$  representa el número de subconjuntos de  $1K$  ( $1K=2^{10}=1024$ ) de palabras en el espacio de direcciones. Por ejemplo, una PDP-11<sup>3,4</sup> tiene un espacio de direcciones de  $64-K$ : palabras de ocho bits o bytes de  $64-K$ . Un byte suele representar un grupo de ocho bits.

La memoria principal es una memoria de acceso aleatorio (o directo). Independientemente del orden en que aparecen las direcciones en el ducto de direcciones, el tiempo que se necesita para introducir los datos en una dirección aleatoria cualquiera es constante. La característica del acceso aleatorio distingue a la memoria principal de los dispositivos de almacenamiento secundario, tales como cintas, discos y tambores magnéticos.

**Procesador I/O (de entrada y salida).** Controla el flujo de información entre dispositivos I/O, la CPU y secciones de memoria de la máquina. Como se muestra en la figura 56-1, el procesador I/O tiene dos canales de comunicación con otros dispositivos (las líneas discontinuas de la figura).

El canal situado entre la memoria y el procesador I/O se llama canal de acceso directo a la memoria (DMA) y es utilizado por dispositivos que pueden transferir datos a grandes velocidades (cercanas a la del reloj del procesador). El canal DMA acelera la operación global de la computadora, ya que la CPU no tiene que intervenir directamente en la transacción, lo que hace posible la ejecución concurrente de programas y la transferencia de datos.

El canal DMA se forma físicamente mediante la habilitación del procesador I/O para que tenga control sobre los ductos de direcciones y de datos durante el tiempo en que la CPU no requiere acceso a la memoria. Esta técnica se denomina robo del ciclo.

En el caso de dispositivos cuyas velocidades de transferencia de datos son bajas con respecto a la velocidad del reloj del procesador, la transferencia de los datos suele ser controlada por la CPU. En este caso el procesador I/O controla la entrada (salida) de datos del (al) dispositivo periférico, almacena los datos temporalmente y manda una señal a la CPU cuando los datos están disponibles (o bien cuando el dispositi-

\* Un bit es una cifra binaria cuyo valor es 0 o 1.

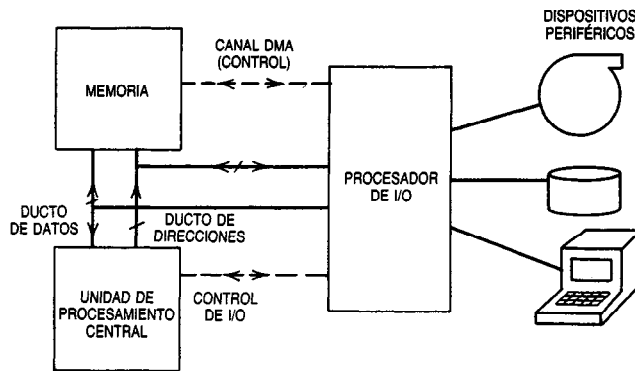


Fig. 56-1. Diagrama de bloques de un sistema de computación general.  
DMA = acceso directo a la memoria; I/O = entrada-salida.

vo periférico está listo para recibir datos nuevos). En este modo de transferencia de datos, el procesador I/O nunca obtiene el control de los ductos de direcciones y de datos.

En ambos modos de transferencia, el procesador I/O interpreta instrucciones especiales para realizar el proceso de manera confiable. Las instrucciones se detallan en la sección 56.3.

## 56.2 REGISTROS DE LA CPU

La CPU contiene un conjunto de localidades de almacenamiento temporal de datos de alta velocidad llamadas registros. Algunos de los registros están dedicados al control, y sólo la unidad de control tiene acceso a ellos. Los registros restantes son registros de uso general, y el programador tiene acceso a ellos. En la explicación que sigue se aclarará esta distinción.

Dentro del conjunto básico de registros de control se deben incluir:

- Contador del programa (PC).
- Registro de direcciones de la memoria (MAR).
- Registro de datos de la memoria (MDR).
- Registro de instrucciones (IR).
- Palabra de estado del programa (PSW).

**PC.** La función del PC consiste en seguir la pista de la instrucción por buscar (capturar) en el "siguiente" ciclo de máquina; por tanto, contiene la dirección de la siguiente instrucción por ejecutar. El PC es modificado dentro del ciclo de búsqueda de la instrucción "actual" mediante la adición de una constante. El número que se agrega al PC es la longitud de una instrucción en palabras. Por tanto, si una instrucción tiene una palabra de longitud, se agrega 1 al PC; si una instrucción tiene dos palabras de largo, se agrega 2; y así sucesivamente.

Si la máquina es de instrucciones con longitud variable, se deberá ejecutar un número de ciclos de búsqueda igual a la longitud de la instrucción en palabras. Este número es determinado por la unidad de control

como resultado de descifrar el código de operación de la instrucción. El número de bits en el PC es el mismo que el del ducto de direcciones.

**MAR.** Funciona como el registro de enlace entre la CPU y el ducto de direcciones. Cuando se logra el acceso a la memoria, la dirección es colocada en el MAR por la unidad de control y ahí permanece hasta que se completa la transacción. El número de bits que hay en el MAR es el mismo que el del ducto de direcciones.

Es oportuno aclarar aquí la diferencia entre el PC y el MAR. Si no es necesario hacer referencia a la memoria durante el ciclo de ejecución de una instrucción, el PC y el MAR sirven al mismo fin. Sin embargo, muchas de las instrucciones de la máquina hacen referencia a la memoria y operan con los datos que están en ella. Como la dirección de los datos suele ser diferente de la dirección de la instrucción siguiente, se necesita el MAR.

**MDR.** La función del MDR consiste en proporcionar un área de almacenamiento temporal (memoria intermedia, acumulador intermedio o buffer) de datos que se intercambian entre la CPU y la memoria. Los datos pueden ser instrucciones (obtenidas en el ciclo de búsqueda) o datos del operando (obtenidos en el ciclo de ejecución). Debido a su conexión directa con el ducto de datos, el MDR contiene el mismo número de bits que él.

**IR.** Es un registro que conserva el código de operación de la instrucción en todo el ciclo de la máquina. El código es empleado por la unidad de control de la CPU para generar las señales apropiadas que controlen la ejecución de la instrucción. La longitud del IR es la longitud en bits del código de operación.

**PSW.** La palabra de estado o condición del programa (PSW) almacena información pertinente sobre el programa que esté ejecutándose. Por ejemplo, al completarse una función de la ALU, se modifica un conjunto de bits llamados códigos (o señales) de condición. Es-

7	6	5	4	3	2	1	0	POSICIÓN DE BIT
1	0	1	1	0	1	0	1	OPERANDO ORIGINAL (A)
1	1	1	1	0	0	0	0	PATRÓN DE BITS DE MÁSCARA (B)
1	0	1	1	0	0	0	0	A Y B

Fig. 56-2. Ilustración de la función AND.

tos bits especifican si el resultado de una operación aritmética fue cero o negativo (o positivo) o si el resultado se desbordó, en el caso de la aritmética de complemento a dos. El programa puede verificar estos bits en las instrucciones siguientes y cambiar en forma condicional su flujo de control según su valor.

Además, el PSW contiene bits que hacen posible que la computadora responda a solicitudes de servicio asincrónicas generadas por dispositivos I/O o condiciones de error internas. Las señales se denominan de interrupción, y se detallan en la sección 56.5.

Los registros restantes son de uso general. Éstos se utilizan para almacenar información temporalmente, y también retienen operandos que participan en operaciones de la ALU. Algunas veces el conjunto de instrucciones de la computadora y el esquema de direccionamiento de la arquitectura restringen el uso de algunos de estos registros. Por ejemplo, en el caso de la DEC PDP-11, que acepta el direccionamiento por pila, se supone que un registro específico retiene la dirección de la pila, mientras que en el sistema IBM 370, que acepta el direccionamiento relativo a la base, cuando menos se deberá utilizar uno de los registros para retener la dirección base (primera) del programa que esté ejecutándose en ese momento.

Si bien en todas las máquinas la información contenida en el registro puede manipularse como datos ordinarios, durante la ejecución de algunas instrucciones los datos se utilizan en forma explícita para deducir una dirección de la memoria. La ventaja de usar registros para retener datos de operaciones es la velocidad. Si no necesitan buscarse operandos en la memoria, el cálculo promedio con dos operandos se acelera en forma sustancial.

### 56.3 TIPOS DE INSTRUCCIONES

En una computadora, las instrucciones pueden clasificarse de manera amplia en cinco categorías:

- Instrucciones aritméticas y de lógica.
- Instrucciones de movimiento de datos.
- Operaciones de datos en bloque.
- Instrucciones de control del programa.
- Instrucciones de I/O.

#### Instrucciones aritméticas y de lógica

Entre ellas se cuentan operaciones binarias y unarias. Las operaciones binarias requieren dos operandos y producen un resultado único. La suma y la resta, y

también la multiplicación y la división, son operaciones estándar en la mayor parte de las máquinas, con la excepción de algunas minicomputadoras y microprocesadores.

Las operaciones de lógica incluidas en la mayor parte de los conjuntos de instrucciones son las operaciones AND, INCLUSIVE OR y EXCLUSIVE OR.

**Operación AND.** Se efectúa bit por bit entre dos operandos y produce un uno lógico solamente si ambos bits son iguales a 1. Un uso importante de la función AND consiste en aislar un subconjunto de bits de un operando más largo. La figura 56-2 ilustra el procedimiento en forma gráfica. Supóngase que se aislaron los bits del 4 al 7 de un operando de ocho bits. Al aplicar la función AND entre el operando original y el patrón de bits 11110000 se fuerza a los bits menos significativos del resultado a ser cero (los bits del 3 al 0); los bits más significativos del resultado son iguales al patrón de bits original de los bits 4 al 7. Si el bit original era un 1 no se altera al aplicarle la operación AND con un 1; si el bit original era un 0, la operación AND retiene el valor en forma similar. Esta operación de filtrado se conoce como enmascaramiento.

**Operación INCLUSIVE OR<sup>3,5</sup> (O incluyente).** Puede utilizarse para detectar si un operando es igual a cero. Si se aplica el INCLUSIVE OR al operando original y un conjunto de ceros, la única ocasión en que el resultado es cero es cuando el operando original contiene ceros; de lo contrario el resultado es el operando original. (La operación INCLUSIVE OR de dos bits es cero sólo si ambos bits son cero.)

**Operación EXCLUSIVE OR<sup>3,5</sup> (O excluyente).** Produce un 1 entre dos bits sólo si ambos bits difieren. Esta operación podría utilizarse para detectar variaciones en los dispositivos I/O. Supóngase que cada bit de un operando corresponde a un indicador de encendido y apagado de una serie de dispositivos, y que los bits se determinan como resultado de leer el estado de los dispositivos (vía el procesador I/O). Si el barrido se efectúa en forma periódica aplicando la operación EXCLUSIVE OR al operando (estado barrido previo) con el nuevo estado barrido, sólo figuran unos en aquellos indicadores de dispositivos que cambiaron. Esto se ilustra en la figura 56-3.

Además de sus funciones primarias, las instrucciones aritméticas y de lógica activan las señales o códigos de

BIT	7	6	5	4	3	2	1	0	
	0	1	1	0	1	0	0	1	ESTADO ANTERIOR
	1	0	1	1	0	0	1	1	NUOVO ESTADO
	1	1	0	1	1	0	1	0	ANTERIOR EXOR NUOVO

Fig. 56-3. Ilustración de la función EXCLUSIVE OR (EXOR, OR excluyente). Los unos indican cambios.

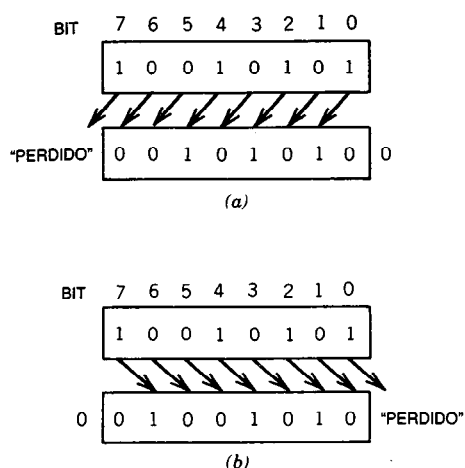


Fig. 56-4. Operaciones de desplazamiento: a) a la izquierda; b) a la derecha.

condición (sec. 56.2). Por tanto, si el sistema aritmético es de complemento a dos, el hardware de operaciones de adición normalmente detectará una condición de desbordamiento, un acarreo desde el bit más significativo, y si el resultado fue igual a cero o fue positivo o negativo. Los códigos de condición se activan en forma correspondiente.

Las operaciones unarias requieren un solo operando y producen un solo resultado.

El cambio del signo de un operando se realiza por medio de la operación de negación (NEG). En la mayor parte de las máquinas esto da origen a que se obtenga el complemento a dos del resultado.

Las operaciones de incremento (INC) y decremento (DEC) son medios de conteo para controlar ciclos del programa, y son eficientes para modificar direcciones contenidas en un registro de índice (sec. 56.4).

También es frecuente el uso de una operación de borrado<sup>3</sup>, en la cual el operando se hace igual a cero.

El desplazamiento y la rotación<sup>6</sup> son otras operaciones unarias incluidas en la mayor parte de las instrucciones. Los desplazamientos pueden realizarse a la derecha o a la izquierda, y ser lógicos o aritméticos.

**Operaciones de desplazamiento.** Un desplazamiento lógico<sup>7</sup> traslada los bits originales del operando una posición (a la izquierda o a la derecha). En un desplazamiento hacia la izquierda se pierde el bit más significativo del operando original, mientras que en un desplazamiento a la derecha se pierde el bit menos significativo. La posición del bit que queda vacante es ocupada con un cero en ambos casos (fig. 56-4). El desplazamiento aritmético hacia la izquierda es idéntico al desplazamiento lógico hacia el mismo lado. Sin embargo, en un desplazamiento aritmético a la derecha, el bit más significativo se trata como un bit de signo, y el bit del signo se conserva en el bit más significativo. El desplazamiento hacia la izquierda corresponde a la multiplicación por 2, y a la derecha corresponde a la división entre 2.

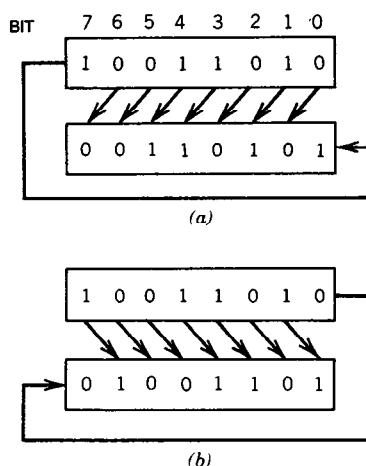


Fig. 56-5. Operaciones de rotación: a) a la izquierda; b) a la derecha.

**Operaciones de rotación.** Son similares a las de desplazamiento, salvo que en una rotación a la izquierda el bit más significativo se convierte en el menos significativo, mientras que en una rotación a la derecha sucede exactamente lo contrario. Esto se ilustra en la figura 56-5.

Las operaciones de rotación constituyen una manera ventajosa de verificar si un bit es 1 o 0 sin pérdida de información. El bit por examinar puede hacerse girar a la izquierda, hacia el bit del signo del operando, y el código de condición correcto se activa como resultado. En algunas arquitecturas, en las instrucciones de rotación y desplazamiento, se incluye el bit de acarreo de la ALU<sup>3,4,8-10</sup>.

#### Instrucciones de movimiento de datos

Dan por resultado la copia de datos desde una localidad de operando a otra. Además del código de operación, estas instrucciones requieren información que identifique los operandos fuente y destino.

En una computadora de uso general, los datos se pueden mover de:

1. Registro a registro.
2. Registro a memoria.
3. Memoria a registro.
4. Memoria a memoria.

La mayor parte de las máquinas modernas también tienen la capacidad de movimiento de datos en bloque. Por ejemplo, los conjuntos de instrucciones de los sistemas IBM 370<sup>6</sup> y Zilog Z-8000<sup>11</sup> contienen movimientos de registro a memoria y de memoria a registro que desplazan todos los operandos del registro hacia la memoria principal o desde ella. Estas funciones se realizan normalmente cuando se interrumpe la operación de una máquina y cuando se regresa de un programa de servicio con interrupciones. Las mismas máquinas tienen asimismo un movimiento de memoria a memoria para un bloque de datos. Estas instrucciones

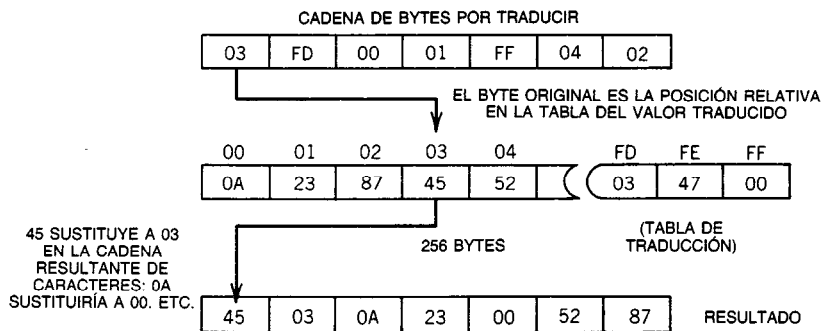


Fig. 56-6. Instrucción de traducción.

requieren, además de la información de la fuente y el destino, el número de bytes que se moverá. La última información debe codificarse como parte de la instrucción.

### Operaciones de datos en bloque

Son aquellas que se efectúan con un conjunto de operandos y no con un solo operando. (Si bien las instrucciones de movimiento de datos en bloque podrían considerarse dentro de esta categoría, se percibió que están conceptualmente mejor situadas en la categoría del movimiento de datos.) Estas operaciones se encuentran cada vez con mayor frecuencia en computadoras recientes, por ejemplo, la Z-8000 y la Intel 8086<sup>11</sup>.

La búsqueda de un conjunto de operandos en la primera presencia de un patrón de bits específico se realiza por medio de una operación de comparación y repetición. Dentro de la instrucción se especifica el operando de prueba, la dirección inicial del conjunto de operandos y el operando que contiene la longitud del conjunto. Esta instrucción es útil, por ejemplo, en una operación de edición, donde la cadena de entrada podría representar una información de texto que se introduce desde un teclado y el objetivo consiste en eliminar caracteres que representen espacios en blanco.

La comparación de dos conjuntos de operandos contenidos en la memoria sobre una base de operando por operando es otra instrucción común de datos en bloque. Esta instrucción es útil cuando se busca en una tabla de símbolos (compuesta de un conjunto de operandos) la presencia o ausencia de un símbolo específico de la tabla. Esta instrucción requiere la dirección inicial de ambos operandos y también la del operando de longitud.

"Traducción" es una instrucción que permite sustituir un conjunto de datos sobre una base de uno a dos, operando por operando, a partir de una tabla de traducción. Esta instrucción opera como sigue (fig. 56-6). Como parte de la instrucción, se especifica la dirección inicial del conjunto de operandos fuente (los datos por traducir), junto con la dirección inicial de la tabla de traducción. Además, se especifica la longitud del operando fuente. Después, cada operando se suma en forma numérica a la dirección inicial de la tabla de

traducción y se reemplaza por el operando que se encontró en la dirección resultante. Este proceso continúa hasta que se sustituye todo el operando fuente.

Una aplicación de esta instrucción sería en un entorno de procesamiento de comunicaciones. Supóngase que a través de una terminal se introduce un mensaje en EBCDIC (*extended binary coded decimal interchange code*, código de intercambio decimal codificado en binario extendido), pero que en la transmisión de mensajes por enlaces de comunicaciones en la red se utiliza ASCII (*american standard code for information interchange*, código estadounidense estándar para el intercambio de información). Por medio de esta instrucción es posible traducir el mensaje en su totalidad.

### Instrucciones de control del programa

Hacen posible que un programa se aparte de la naturaleza secuencial inherente del ciclo de máquina de la computadora. En otras palabras, se pueden pasar por alto secciones de instrucciones como resultado de la activación de un código de condición o como resultado directo del diseño del programa. La instrucción que corresponde al primer caso es la instrucción de salto adicional, mientras que en el segundo caso se utilizaría una instrucción de salto incondicional. Ambos tipos de instrucciones de salto requieren la siguiente instrucción. La ejecución se realiza simplemente colocando el PC en esa dirección. La instrucción de salto condicional deberá contener además la posición del código de condición correspondiente a la ejecución del salto. En caso de que no se cumplan las condiciones, se ejecuta la siguiente instrucción en la secuencia.

En la clase de instrucciones de control del programa también se incluyen las instrucciones de "llamada" (*call*) a subrutinas. Estas pueden ser condicionales o incondicionales. Su operación es similar a la de las instrucciones de salto, con una diferencia importante: la dirección de la siguiente instrucción en la secuencia se guarda en la memoria o en registros de la máquina.

La dirección a la cual se hace la transferencia por medio de una instrucción de llamada es la primera de la subrutina. Después se procede a ejecutar la subrutina. La última instrucción de la subrutina es una instrucción "de regreso" (*return*), que se ejecuta recuperando la dirección que guarda la "llamada" y colocándola en el PC.

Las subrutinas se utilizan por dos razones principales en la estructuración de un programa. Primero, si las mismas instrucciones figuran en forma repetida en diferentes áreas de un programa, se pueden realizar en una subrutina para ahorrar espacio de memoria. Segundo, un programa complejo puede dividirse conceptualmente en una serie de unidades de menor tamaño, cada una de ellas codificada en forma adecuada en una subrutina. Por tanto, el programa original podría constar principalmente de llamadas a subrutina. Una ubicación común del almacenamiento de direcciones de regreso es una *pila*<sup>4,12,13</sup> (sec. 56.4), una estructura de datos donde el último en entrar es el primero en salir, controlada por la CPU. Si se aplica este método, la dirección de regreso es colocada en la parte superior de la pila por la instrucción "call", y retirada de la cima de la pila por la instrucción "return".

#### Instrucciones I/O<sup>2,6,8,14-16</sup>

No se pueden generalizar fácilmente, ya que dependen en gran medida del sistema de computación de interés y también de los dispositivos específicos a los cuales se vincula el sistema. Sin embargo, se puede hacer una clasificación aproximada en base a la estructura de ductos del sistema de computación.

En algunas CPU, los ductos de datos y direcciones se multicanalizan en el tiempo entre todo el equipo periférico conectado. En particular, todos<sup>2,4</sup> los modelos de la PDP-11 poseen una arquitectura de este tipo (el ducto del sistema se llama UNIBUS). Además, la mayor parte de los microprocesadores se construyen con una disposición de ductos similar<sup>11</sup>. La CPU no genera señales de control especiales para los dispositivos que distinguen los datos limitados para la memoria de aquellos otros limitados para un controlador de un dispositivo periférico. La CPU tampoco distingue entre datos que se leen de la memoria o de un dispositivo periférico.

Desde el punto de vista de la programación, para el acceso a la memoria o a un periférico simplemente se requiere el mismo conjunto de instrucciones. Los sistemas que operan en la forma descrita se denominan sistemas I/O mapeados por memoria.

La programación de un dispositivo en sistemas I/O mapeados por memoria requiere el conocimiento del dispositivo y sus características, aunque no se necesitan instrucciones especiales. El dispositivo se caracteriza como un conjunto de localidades de la memoria que se dividen en dos subcategorías: un conjunto de registros de estado y control, y un conjunto de registros de información.

**Registros de estado y control.** Suelen contener información acerca del estado del dispositivo, por ejemplo, si está inactivo u ocupado, si se le está realizando mantenimiento, etc. En estos registros también se almacena información de control, tal como el tipo de paridad (par o impar) y la velocidad de transmisión de datos. La información contenida en los registros de estado y control se utiliza principalmente para proporcionar una imagen global del hardware cuando éste se programa.

MSB	BYTES																		LSB
	ALTO									BAJO									
15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00				
CONTROL Y ESTADO (CSR)	RO	RW	RO	RW	NO USADO	RO	TLINE C	RO	TLINE B	RO	TLINE A	RO	RW	RW	RW	NO USADO			
	TRDY	TIE	SA	SAE				RDONE	RIE	MSE	CLR	MAINT				NO USADO			
RECEPTOR DE ALMACENAMIENTO INTERMEDIO (REUF)	RO	RO	RO	RO	RO	RX LINE C	RO	RO	RBUF D6	RO	RBUF D4	RO	RBUF D3	RO	RBUF D2	RO			
	DATA VALID	OVRN	FRAM ERR	PAR ERR	NO USADO	RX LINE B	RX LINE A	RBUF D7	RBUF D5	RBUF D4	RBUF D3	RBUF D2	RBUF D1	RBUF D0					
PARÁMETRO DE LÍNEA (LPR)	NO USADO	NO USADO	NO USADO	WO RX ON	WO	FREQ D	FREQ C	FREQ B	FREQ A	WO	STOP CODE	CHAR LGTH B	WO	WO	WO	WO			
								ODD PAR	ENAB										

Fig. 56-7. Esquema de registros de estado y de almacenamiento intermedio de la DZ-11. Tomado de Digital Equipment Corp.<sup>17</sup>, con permiso.

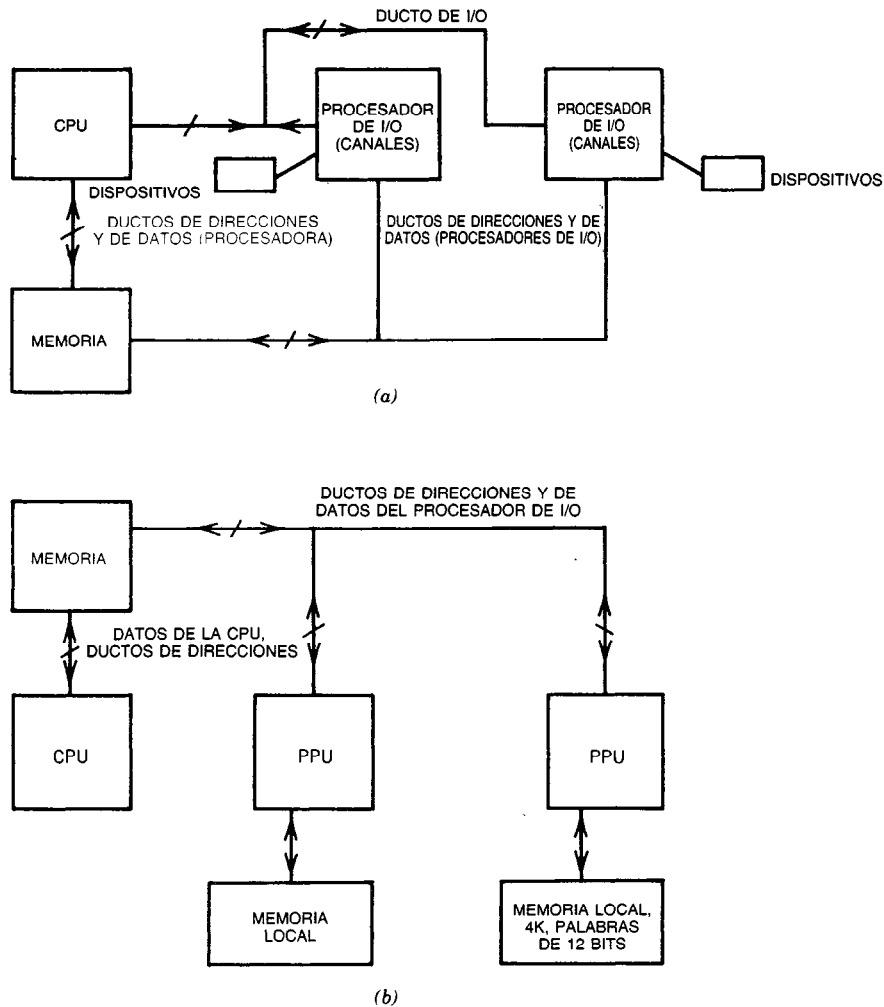


Fig. 56-8. Estructuras de ductos de sistemas de computación: a) IBM 370; b) CDC 6600. CPU = unidad de procesamiento central; I/O = entrada-salida; PPU = unidad de procesamiento periférica.

**Registros de información.** Constituyen una memoria intermedia para la información que se transfiere entre la CPU y el periférico. En el caso de un dispositivo que transfiere datos sobre una base de carácter por carácter, suele haber sólo dos registros, uno que retiene datos transferidos de la CPU al dispositivo, y otro que retiene datos transferidos del dispositivo a la CPU. Si la I/O programada se realiza en un dispositivo unidireccional (sólo transmite o sólo recibe), entonces únicamente se necesita un registro.

En contraste con lo que ocurre en el caso de la I/O controlada por el programa, un dispositivo DMA requiere por lo general más registros de información. En este caso la CPU deberá proporcionar una dirección inicial contenida en la memoria principal para los datos y el número de bytes por transferir. Estos dos registros, además de las memorias intermedias de datos, son necesarios en los dispositivos DMA. La figura 56-7

ayuda a ilustrar estos conceptos de manera más concreta. La figura muestra el esquema del registro de estado y del registro intermedio (o de almacenamiento intermedio) del receptor de una interfaz DZ-11 que conecta la CPU de la PDP-11 a una línea de comunicaciones asincrónica, vía el UNIBUS. La línea puede emplearse para transmisión y recepción, pero aquí solamente se considera el receptor. En la referencia bibliográfica 17 se presentan detalles del transmisor.

Los registros de estado e intermedio del receptor están ubicados en las direcciones 174000 y 174004, respectivamente. Éstas son direcciones estándar reservadas para la unidad. La función del registro intermedio del receptor consiste en almacenar temporalmente los caracteres que se leen de la línea. En el registro de estado existe un conjunto de bits (15-12) dedicados a la detección de diversas condiciones de error que ocurren dentro de la unidad misma. Por ejemplo, los bits 14 y 13 se activan cuando ocurre un evento inesperado den-

CÓDIGO DE OPERACIÓN	CAMPO DE DATOS	CAMPO DE DATOS	CAMPO DE DATOS
	0	1	N

Fig. 56-9. Formato de instrucción general.

tro del dispositivo (p. ej., si la portadora del modem se apaga en la línea, se activa el bit 14), mientras que se activa el bit 12 si la memoria intermedia del receptor no ha sido leída por el programa y se introduce otro carácter al periférico. Los bits 10 y 9 hacen posible que el programa seleccione la longitud de los caracteres vía software, mientras que la velocidad de transmisión digital del dispositivo se puede programar en forma parecida mediante la activación apropiada de los bits 3 y 4. El bit 6, cuando se activa, hace posible que el dispositivo interrumpa a la CPU si se recibe un carácter, mientras que el bit 7 indica si la memoria intermedia del receptor contiene un carácter. Los bits restantes son bits de control que se emplean cuando el DZ-11 se conecta a un módem.

El método alternativo para realizar la I/O consiste en proporcionar un ducto aparte y capacidad de almacenamiento de programas fuera de la CPU. En la serie de computadoras IBM 370<sup>6,16</sup> y en las computadoras CDC 6600/7600<sup>1</sup> se emplea este método. Las estructuras de ductos físicas de estos sistemas se ilustran en la figura 56-8.

La ventaja principal de un ducto separado sobre la estructura de ductos multicanalizados es el mayor rendimiento, ya que con ductos de I/O y de memoria separados existen dos procesadores que realmente trabajan en paralelo. El precio de esta mejora es un costo más elevado debido a la estructura de ductos separados y, además, la necesidad de que los programadores de sistemas aprendan otro lenguaje de programación (el procesador de I/O es una máquina de programa almacenado).

Dos ejemplos de máquinas con este tipo de estructura I/O son las series de computadoras IBM 370 y CDC Cyber 70. En la serie 370 se emplea el esquema I/O de amo y esclavo, donde la CPU tiene un control de I/O específico a través de instrucciones para accionar y detener el procesador de I/O. En la serie CDC Cyber 70 se utiliza un esquema de I/O "de acoplamiento laxo". La CPU pide la I/O, vía una localidad de la memoria común. En este caso, la CPU no tiene comandos I/O explícitos, y los procesadores I/O (unidades de procesamiento periféricas) trabajan de manera asincrónica e independiente con respecto a la CPU.

Las dos máquinas difieren en la estructura de sus procesadores I/O. En el sistema 370, el procesador I/O<sup>6,16</sup> comparte la memoria principal con la CPU y tiene un repertorio de instrucciones más bien limitado. En el CDC, cada PPU tiene una memoria de programa distinta de 4 096 palabras y un conjunto de instrucciones que consta de 66 operaciones diferentes. La única área que se comparte<sup>1</sup> en la memoria principal se utiliza para transmitir mensajes entre los procesadores.

#### 56.4 FORMATOS DE INSTRUCCIONES Y MODOS DE DIRECCIONAMIENTO

Para implementar un conjunto dado de instrucciones en una computadora, es necesario elegir el esquema de los bits de las instrucciones. Dicho esquema se denomina formato de la instrucción. Por lo general, una instrucción consta de un código de operación y otros datos pertinentes para completar con éxito la operación. Por ejemplo, si el código de operación especifica una operación aritmética, la localidad de los operandos se deberá especificar en forma explícita (o implícita), así como la localidad del operando resultante. En el caso de operaciones con datos en bloque (p. ej., una operación de movimiento de datos en bloque de una memoria a otra), en la instrucción deberán incluirse la dirección inicial de ambos bloques y la cuenta total de las palabras que se cambiarán de sitio.

La forma más general de una instrucción se presenta en la figura 56-9. La instrucción se divide en una serie de campos. El primer campo se conoce como código de operación, y los campos restantes se llaman campos de datos. Estos últimos se analizan al final de esta sección; aquí el interés se concentra en el código de operación.

El objetivo principal del código de operación es especificar una acción para la computadora. Además, el código contiene información referente a la longitud de la instrucción. La mayor parte de las computadoras modernas tienen instrucciones de longitud variable para máxima flexibilidad y mínima utilización de la memoria. Instrucciones más complejas (tipos de datos en bloque) requieren las mayores longitudes, mientras que las instrucciones simples (transferencia de datos de un registro de uso general a otro) requieren las longitudes más cortas. La única restricción que se impone a las longitudes de instrucciones es que sean un múltiplo entero de la longitud de palabra de la máquina. En la figura 56-10 se presentan los diversos formatos de instrucciones de la computadora IBM 370. Existen cinco formatos en los cuales se puede poner una instrucción.

**Formato RR (registro-registro).** Este formato tiene tres campos. El código de operación especifica la acción, y los campos restantes designan uno de 16 registros de uso general como operandos fuente y destino. La longitud de esta instrucción es de 16 bits.

Las instrucciones de 32 bits se clasifican como instrucciones RX (registro-memoria), SI (almacenamiento-inmediato) y RS (registro-almacenamiento).

**Formato RX.** En éste, el campo R1 especifica un registro de uso general como un operando, mientras que los



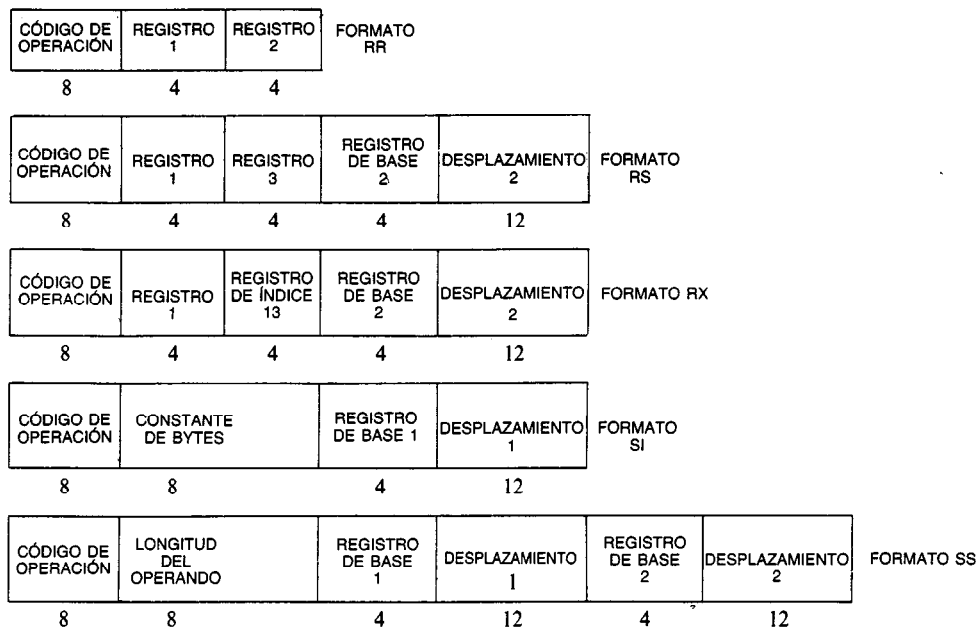


Fig. 56-10. Formatos de instrucciones, sistema 370. RR = registro-registro; RS = registro-almacenamiento; RX = registro-memoria; SI = almacenamiento-uso inmediato; SS = almacenamiento-almacenamiento.

campos X (índice), B (base) y D (desplazamiento) especifican una dirección de la memoria mediante el uso del esquema de direccionamiento de índice-base-desplazamiento común a estas computadoras (que se analiza más adelante).

**Formato SI.** Especifica una constante de ocho bits como un operando (campo de constante) y los campos B y D especifican la dirección de otro operando.

**Formato RS.** Es similar al RX, salvo que en este caso el campo R3 se utiliza como una tercera dirección de operando.

**Formato SS de 48 bits (almacenamiento-almacenamiento).** En este caso los campos B y D especifican las direcciones iniciales de dos operandos de bloque y el campo de longitud, donde las longitudes de los bloques participan en la operación.

En las máquinas 370, el código de operación se fija en ocho bits sin importar la longitud de la instrucción.

En contraste con las máquinas que tienen un código de operación fijo, algunas tienen códigos de operación de longitud variable. Estos códigos se denominan códigos de operación de expansión.

La mayor parte de las computadoras de uso general aceptan conjuntos de instrucciones que contienen cero, uno, dos y tres campos de dirección. Por ejemplo, una arquitectura de máquina que acepte el direccionamiento por pila puede contener una operación de introducción (cero direcciones), una operación de complemento (instrucción de una dirección), una

operación de adición (dos direcciones) y una operación de movimiento de datos (tres direcciones). Considérese esta arquitectura y supóngase que el número de instrucciones de tres direcciones de la máquina es 10, el de instrucciones de dos direcciones es 12, el de instrucciones de una dirección es 13 y el número de instrucciones de cero direcciones es 10. Supóngase además que la información de los campos de direcciones tiene cuatro bits de longitud. (Si bien a simple vista la limitación de direcciones parece restrictiva, si la máquina retiene toda su información de direcciones en uno de los 16 registros de uso general, esto se justifica.) El esquema del código de operación de expansión de esta máquina se ilustra en la figura 56-11. A las instrucciones que requieren el número máximo de campos de direcciones se les asignan códigos de operación de cuatro bits; el código 1111 está reservado como prefijo de todos los códigos de operación de dos direcciones. El doble prefijo de ocho bits 1111 1111 se asigna a todos los códigos de operación de una sola dirección, y por último, el prefijo de 12 bits 1111 1111 1111 se asigna a todos los códigos de operación de cero direcciones. Debido a la simetría del ejemplo, el tamaño de instrucción básico es de 16 bits, e independiente de los modos de direccionamiento de la máquina. La transición hacia cada nuevo nivel del código de operación, como se describe en el ejemplo, hace que se disponga solamente de 15 códigos de operación en cada nivel\*, ya que cada uno de ellos se define como prefijado por un número de unos. Supóngase que la máquina requiere 30 instrucciones de dos direcciones (códigos de

\*Excepto en el nivel de cero direcciones.

4 BITS (CÓDIGOS DE OPERACIÓN DE TRES DIRECCIONES)				XXXX ≠ 1111 15 CÓDIGOS DE OPERACIÓN
XXXX	A1	A2	A3	
8 BITS (CÓDIGOS DE OPERACIÓN DE DOS DIRECCIONES)				XXXX ≠ 1111 15 CÓDIGOS DE OPERACIÓN
1111	XXXX	A1	A2	
12 BITS (CÓDIGOS DE OPERACIÓN DE UNA DIRECCIÓN)				XXXX ≠ 1111 15 CÓDIGOS DE OPERACIÓN
1111	1111	XXXX	A1	
16 BITS (CÓDIGOS DE OPERACIÓN DE CERO DIRECCIONES)				XXXX ≠ 1111 16 CÓDIGOS DE OPERACIÓN
1111	1111	1111	XXXX	

Fig. 56-11. Expansión de los códigos de operación (op).

operación de ocho bits). Es claro que el esquema que se ilustra no funcionaría. Un método para resolver este problema se ilustra en la figura 56-12. En este caso se utilizan dos prefijos de cuatro bits, 1110 y 1111, para realizar la transición de códigos de operación de cuatro a ocho bits, lo que supone una reducción del número de instrucciones de tres direcciones de 15 a 14. El prefijo extra permite definir otros 16 códigos de operación de ocho bits. En la tabla 56-1 se ilustra el concepto de código de operación de expansión como se implementa en un procesador específico, el PDP-11/70.

Una alternativa al método de los prefijos de variar la longitud de los códigos de operación consiste en asignar un campo contenido en la instrucción que cuando se decodifique especifique la longitud. Este método hace posible la decodificación completa del número de bits en cada nivel, lo que incrementa en forma considerable el número de códigos de operación disponibles para su uso. Por ejemplo, si se asignaran dos bits para definir la longitud de las instrucciones del ejemplo anterior, se dispondría de 16 códigos de tres direcciones, 256 ( $2^8$ ) códigos de dos direcciones, 4 096 ( $2^{12}$ ) códigos de una dirección y 65 536 ( $2^{16}$ ) códigos de cero direcciones. El costo sería dos bits por palabra de instrucción (fig. 56-13).

#### Modos de direccionamiento de una computadora

El término se refiere a los métodos que se aplican en la máquina para generar una dirección del operando. Todas las instrucciones requieren cuando menos un operando para su ejecución; muchas requieren dos o más. Los operandos se almacenan en un registro de

uso general de la CPU o en la memoria. Si el operando está en un registro de uso general de una CPU que tiene  $k$  registros de este tipo, entonces se necesita un mínimo de  $\log_2 k$  bits para especificar la localidad del registro dentro de la CPU. Por ejemplo, volviendo a la figura 56-10, las especificaciones RR, RX y RS implican un operando almacenado en un registro de la CPU. En este caso  $k=16$  y se necesitan cuatro bits para codificar la información de direcciones. Sin embargo, en el caso de un operando de la memoria se deberá especificar una dirección completa. Esta dirección se denomina *dirección efectiva* del operando y puede especificarse en varias formas. En lo que resta de este análisis, se supondrá que la longitud de una dirección efectiva se expresa en bits.

**Direccionamiento directo<sup>5</sup>.** Se emplea para generar direcciones de la memoria y consiste en incluirlas en la instrucción. No se necesitan cálculos en la CPU para determinar una dirección. Como parte de la secuencia de ejecución, la dirección se lee en la CPU (durante el ciclo de búsqueda), y después el operando de esa dirección se lee en la CPU para que participe en la acción que dicta la instrucción. En una computadora que acepta el direccionamiento directo, se requieren dos accesos a la memoria (búsqueda de la instrucción, búsqueda del operando) para ejecutar la instrucción.

Una desventaja importante del direccionamiento directo es el gran número de bits que se requiere cuando en una instrucción se incluye más de un operando. Por ejemplo, en una instrucción de movimiento de datos en bloque en el sistema 370<sup>6</sup>, se requieren dos operandos de dirección. Una dirección especifica la fuente y otra el destino. Como las direcciones tienen 24 bits en esta máquina, se requeriría un total de 48 bits para ambos operandos, así como ocho bits para el código de operación y bits extra para la longitud del bloque. Cuando se deshabilita el direccionamiento directo, esta instrucción se ejecuta mediante el uso de sólo 32 bits de información de direcciones. El que sigue es un ejemplo de la abrumadora diferencia en longitudes de instrucción entre el direccionamiento directo y el direccionamiento de registros. El formato de la instrucción ADD de una PDP-11<sup>4,18</sup> con direccionamiento directo se ilustra en la figura 56-14. Obsérvese que la instrucción tiene 6 bytes de longitud. La dirección fuente de 16 bits sigue a los dos primeros bytes del código de operación, y la dirección destino sigue a la dirección fuente. Si los operandos se almacenan en registros, la instrucción requiere únicamente dos bytes, y los registros destino en efecto son parte del código de operación.

XXXX	A1	A2	A3	XXXX ≠ 1111 O 1110 14 CÓDIGOS DE OPERACIÓN
1111 O BIEN 1110	XXXX	A2	A3	
				XXXX ≠ 1111 30 CÓDIGOS DE OPERACIÓN

Fig. 56-12. Incremento del número de códigos de operación (op) disponibles (doble prefijo).

Tabla 56-1. Códigos de operación de la PDP-11

**Grupo 1.** Códigos de operación de cuatro bits. Formato<sup>a</sup>: xxxssssssdddddxxxx

0000	—	Véanse los grupos 2 y 3
0001	MOV	Mover
0010	CMP	Comparar
0011	BIT	Prueba de bits
0100	BIC	Borrado de bits
0101	BIS	Colocación (activación de bits)
0110	ADD	Sumar
0111	—	Véase el grupo 10
1000	—	Véanse los grupos 11, 12 y 13
1001	MOVB	Mover bytes
1010	CMPB	Comparar bytes
1011	BITB	Byte de prueba de bits
1100	BICB	Byte de borrado de bits
1101	BISB	Byte de colocación de bits
1110	SUB	Restar
1111	—	(Instrucciones de punto flotante)

**Grupo 2.** Códigos de operación de ocho bits. Formato: 00000xxxkkkkkkkkxxx

000	—	Véase el grupo 4
001	BR	Ramificación
010	BNE	Ramificación no igual
011	BEQ	Ramificación igual
100	BGE	Ramificación mayor o igual que
101	BLT	Ramificación menor que
110	BGT	Ramificación mayor que
111	BLE	Ramificación menor o igual que

**Grupo 3.** Códigos de operación de siete bits. Formato: 00001xxrrrdddddxx

00	JSR	Salto a la subrutina
01	—	Véase el grupo 5
10	—	Véase el grupo 6
11	—	Repuesto

**Grupo 4.** Códigos de operación de 10 bits. Formato: 00000000xxdddddxx

00	—	Véase el grupo 7
01	JMP	Salto
10	—	Véase el grupo 8
11	SWAB	Intercambiar bytes

Tabla 56-1. Códigos de operación de la PDP-11 (continuación)

**Grupo 5.** Códigos de operación de 10 bits. Formato: 0000101xxxxdddd

xxx

000	CLR	Borrar
001	COM	Complementar (complemento a uno)
010	INC	Incrementar
011	DEC	Disminuir
100	NEG	Negar (complemento a dos)
101	ADC	Sumar acarreo
110	SBC	Restar acarreo
111	TST	Prueba

**Grupo 6.** Códigos de operación de 10 bits. Formato: 0000110xxxxdddd

xxx

000	ROR	Girar a la derecha un bit
001	ROL	Girar a la izquierda un bit
010	ASR	Desplazamiento aritmético a la derecha de un bit
011	ASL	Desplazamiento aritmético a la izquierda de un bit
100	(MARK)	Limpiar la pila. ddddd = cuenta
101	(MFPI)	Mover desde el espacio de instrucción anterior
110	(MTPI)	Mover hacia el espacio de instrucción anterior
111	(SXT)	Extensión del signo

**Grupo 7.** Códigos de operación de 12 bits. Formato: 000000000xxccc

xx

00	—	Véase el grupo 9
01	—	Repuesto
10	CCC	Borrar códigos de condición
11	SCC	Colocar códigos de condición

**Grupo 8.** Códigos de operación de 13 bits. Formato: 0000000010xxxxrr

xxx

000	RTS	Regreso desde la subrutina
011	(SPL)	Fijar el nivel de prioridad

**Grupo 9.** Códigos de operación de 16 bits. Formato: 000000000000xxx

xxx

000	HALT	Suspender
001	WAIT	Esperar
010	RTI	Regreso desde interrupción
011	BPT	Pausa o interrupción
100	IOT	Trampa de I/O
101	RESET	Reiniciar
110	(RTT)	Regreso de trampa

Tabla 56-1. Códigos de operación de la PDP-11 (continuación)

**Grupo 10.** Códigos de operación de siete bits. Formato: 011xxxrrrrdddddxxx

000	(MUL)	Multiplicar
001	(DIV)	Dividir
010	(ASH)	Desplazamiento aritmético
011	(ASHC)	Desplazamiento aritmético combinado
100	(XOR)	OR excluyente
111	(SOB)	Restar uno y ramificar

**Grupo 11.** Códigos de operación de ocho bits. Formato: 10000xxxkkkkkkkkxxx

000	BPL	Ramificación en el signo más
001	BMI	Ramificación en el signo menos
010	BHI	Ramificación en estado alto
011	BLOS	Ramificación en estado bajo o el mismo
100	BVC	Ramificación al eliminarse el desbordamiento
101	BVS	Ramificación al fijarse el desbordamiento
110	BCC	Ramificación al eliminarse el acarreo
111	BCS	Ramificación al fijarse el acarreo

**Grupo 12.** Códigos de operación de 10 bits. Formato: 1000101xxxdddddxxx

000	CLRB	Borrar bytes
001	COMB	Complementar bytes
010	INCB	Incrementar bytes
011	DECB	Disminuir bytes
100	NEGB	Negar bytes
101	ADCB	Sumar el byte de acarreo
110	SBCB	Restar el byte de acarreo
111	TSTB	Probar el byte

**Grupo 13.** Códigos de operación de 10 bits. Formato: 1000110xxxdddddxxx

000	RORB	Girar el byte a la derecha de un bit
001	ROLB	Girar el byte a la izquierda de un bit
010	ASRB	Desplazar el byte a la derecha de un bit
011	ASLB	Desplazar el byte a la izquierda de un bit
100	—	Repuesto
101	(MFPD)	Mover desde el espacio de datos anterior
110	(MTPD)	Mover hacia el espacio de datos anterior
111	—	Repuesto

Fuente: A. S. Tanenbaum<sup>1</sup>, reproducido con permiso.<sup>a</sup> ssssss especifica una fuente; dddddd especifica un destino; rrr especifica un registro; x especifica bits de los códigos de operación; kkkkkkkk especifica un valor de compensación o una constante; cccc especifica bits del código de condición.

L	CO + CAMPOS DE DATOS
2	16
00 ⇒ CÓDIGOS DE OPERACIÓN DE 4 BITS DE TRES DIRECCIONES, 16 INSTRUCCIONES	
01 ⇒ CÓDIGOS DE OPERACIÓN DE 8 BITS DE DOS DIRECCIONES, 256 INSTRUCCIONES	
10 ⇒ CÓDIGOS DE OPERACIÓN DE 12 BITS DE UNA DIRECCIÓN, 4 096 INSTRUCCIONES	
11 ⇒ CÓDIGOS DE OPERACIÓN DE 16 BITS DE CERO DIRECCIONES, 65 536 INSTRUCCIONES	

**Fig. 56-13.** Incremento del número de códigos de operación (op) de longitud variable (campo de longitud). L = longitud de la instrucción; CO = códigos de operación.

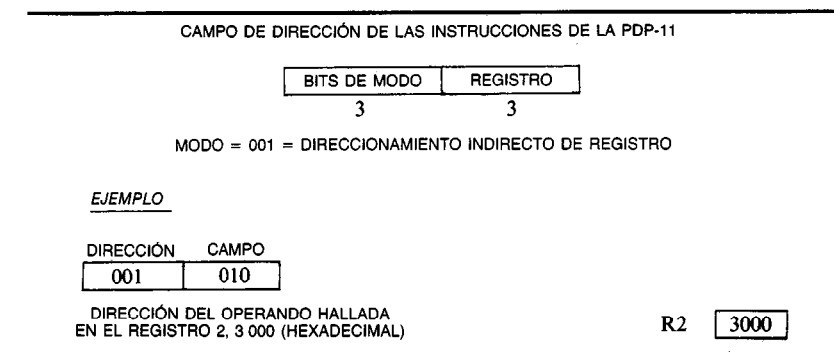
**Direccionamiento indirecto<sup>18</sup> de registros.** Este otro método de direccionamiento consiste en utilizar un registro de uso general de la CPU para almacenar la información. La figura 56-15 muestra un ejemplo. En este caso la dirección del operando está contenida en un registro de uso general. Para especificar esta forma de direccionamiento, la instrucción requiere tres campos: el código de operación, un campo que especifique que el direccionamiento es indirecto, y un campo que especifique el registro. Como el número de registros de uso general es mucho menor que el número de localidades de la memoria, la longitud de la instrucción (para el direccionamiento directo) es corta. El precio que se paga es que la instrucción requiere otro campo de información. Además, el número de accesos de la memoria para la misma operación por direccionamiento indirecto de registros es menor que cuando se utiliza el direccionamiento directo. A fin de poder operar con este método de direccionamiento, los registros de uso general contenidos en la máquina deberán ser lo suficientemente largos para contener una dirección completa.

**Direccionamiento indizado.** Ésta es una extensión de la técnica de direccionamiento indirecto de registros. En una arquitectura de direccionamiento indizado, una dirección de operando se determina sumando una constante almacenada en la instrucción al contenido de un registro para generar una dirección. (Si la constante es igual a cero, el efecto es el mismo que con el direccionamiento indirecto de registros.) En la figura 56-16 se presenta un ejemplo de esta forma de direccionamiento para la PDP-11. El operando fuente de la instrucción está en un registro (R0), mientras que la compensación (constante) a 16 bits se suma al registro R1 para generar la dirección efectiva de 16 bits. La técnica de indización es útil cuando se generan direcciones de memoria consecutivas, como en un programa que realiza operaciones repetitivas sobre un conjunto de operandos, una después de la otra. En este caso la instrucción repetida almacenaría la dirección inicial del conjunto de operandos, y en cada iteración se ajustaría el registro de índice para modificar de manera apropiada la dirección. En la figura 56-17 se ilustra un ejemplo de la técnica. En este caso cada operando se direcciona en forma secuencial sumando 1 al registro de índice en cada iteración del ciclo. Algunas máquinas cuentan con un mecanismo para incrementar en forma automática<sup>1,11</sup> el registro de índice. Esta característica recibe el nombre de autoindización y reduce un programa iterativo en un paso. Cuando se implementa en una máquina, el incremento puede realizarse antes de obtener la dirección efectiva (preindización) o después (posindización).

Una familia de máquinas en la que se emplea la técnica del direccionamiento indizado exclusivamente para determinar direcciones son los sistemas IBM 360/370<sup>6,16</sup>. La dirección efectiva se genera sumando el contenido de uno de los registros de uso general (llamado registro base) a un valor de compensación a

CO + INFORMACIÓN DE MODO	DIRECCIÓN FUENTE	DIRECCIÓN DESTINO
16	16	16

**Fig. 56-14.** Ejemplo de direccionamiento directo. CO = código de operación.



**Fig. 56-15.** Direccionamiento indirecto de registros.

MNEMOTÉCNICO PDP-11

MOV R0, 10 · (R1)

FUENTE DESTINO

INSTRUCCIONES EN LENGUAJE  
DE MÁQUINA (32 BITS)

4	3	3	3	3
0001	000	000	110	001
	0	0	1	0

INSTRUCCIÓN (BINARIA)  
COMPENSACIÓN (HEXADECIMAL)

MODO = 110 ⇒ INDIZADO

R0 35 (HEXADECIMAL)

R1 A 0 1 2 (HEXADECIMAL)

Fig. 56-16. Direccionamiento indizado. El contenido de R0 se mueve hacia la localidad específica por el contenido de R1 más 10 (A022). R1 no se altera.

```

MOV # 10., R0
CLR R1
LOOP: INCB TABLE (R1); SUMAR 1 AL BYTE DE LA TABLA + c(R1)
      INC R1; SUMAR 1 AL REGISTRO DE ÍNDICE
      DEC R0
      BNE LOOP; REGRESAR SI LA CUENTA NO ES CERO
    
```

Fig. 56-17. Direccionamiento indizado, programa de ejemplo. Este segmento de programa de la PDP-11 incrementa los 10 primeros bytes de las localidades TABLE + 0 hasta TABLE + 9 mediante la modificación de direcciones, utilizando el registro de índice R1. R0 es el contador de ciclos.

12 bits. Según el tipo de instrucción, se puede sumar el contenido de otro registro al registro base y el valor de compensación a fin de formar la dirección final. Esta técnica se ilustra en la figura 56-18. Ésta hace posible que los programas contenidos en el sistema sean cambiados de una ubicación a otra sin modificar el código de la máquina.

Obsérvese que si la primera instrucción del programa carga la dirección inicial del programa en el registro base, no se necesita volver a calcular la información de direcciones contenida en cada instrucción, ya que es un valor de compensación relativo al registro base. Al momento del ensamblaje, los valores de compensación se determinan sin dificultad para la dirección inicial cero.

El código que posee este atributo recibe el nombre de código independiente de la posición<sup>1,18</sup> o reubicable. Este tipo de código se adapta bien a un entorno de multiprogramación, en la cual la memoria se comparte entre muchos usuarios.

En arquitecturas de computadora modernas, el concepto de la reubicación se extendió definiendo el registro base como el contador del programa (PC)<sup>4</sup>. Puesto que el PC contiene la dirección de la siguiente instrucción si los valores de compensación se determinan respecto al PC, se genera automáticamente un código reubicable. Esta forma de direccionamiento suele denominarse direccionamiento relativo, y se ilustra en la figura 56-19 para la máquina PDP-11.

El direccionamiento indirecto de la memoria es conceptualmente similar al direccionamiento indirecto de registros (la técnica se conoce asimismo como direccionamiento diferido). El direccionamiento indirecto

se describe mejor por medio de un ejemplo. Considérese la instrucción que se representa en la figura 56-20. La interpretación de la instrucción es "súmese a R0 el operando que se halla en la dirección que especifica el operando de A, y después almacénese el resultado en R0". Obsérvese que los datos de la dirección A se interpretan como una dirección; el operando que se suma a R0 se encuentra en esa dirección. Esta técnica suele aplicarse en máquinas con un número limitado de registros de uso general en sustitución del direccionamiento indirecto de registros. Por ejemplo, considérese la PDP-8<sup>2</sup>. Esta máquina tiene sólo un acumulador, y todas las instrucciones tienen 12 bits de longitud. Excluyendo el código de operación, se utilizan nueve bits de la instrucción para efectuar el direccionamiento. Como una dirección efectiva tiene 12 bits de largo, debe emplearse el direccionamiento indirecto para tener acceso a todo el espacio de direcciones. Los nueve bits de dirección se utilizan de la manera siguiente. La memoria se divide lógicamente en la máquina en 32 páginas de 128 palabras (fig. 56-21). Dos de los nueve bits de dirección los utiliza el hardware como bits de control. El bit 3 de la instrucción se utiliza como un indicador de páginas. Si el bit es 0, la dirección efectiva se determina anexando los siete bits restantes de la instrucción a cinco ceros iniciales, que especifican una dirección en la página 0. Si el bit es un 1, la dirección efectiva se obtiene anexando los siete bits a los cinco bits más significativos del contador del programa (la página actual que el programa está ejecutando). El bit 4 es el indirecto. Si es un 1, los datos de la dirección efectiva se utilizan como una dirección de 12 bits, y en caso contrario se usan como un operando.

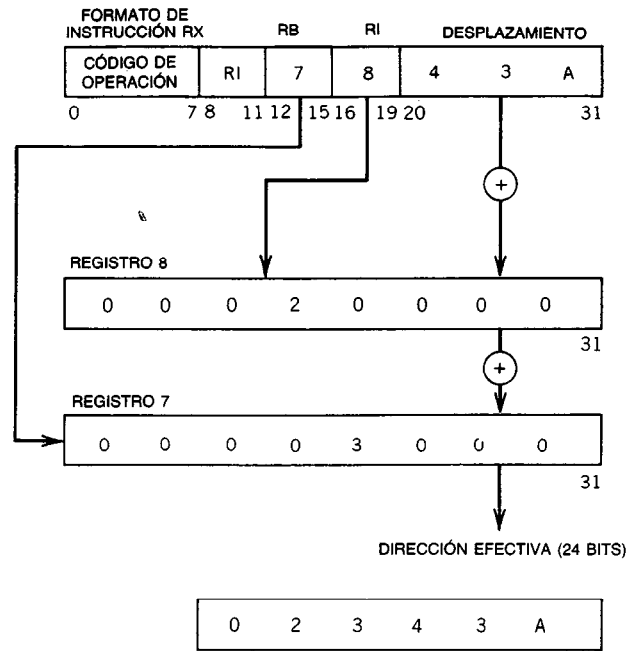


Fig. 56-18. Direccionamiento indizado por la base, IBM 370. Todos los valores están en hexadecimal. RB = registro de base; RI = registro de índice.

## 56.5 MODOS DE OPERACIÓN

Los grandes sistemas de computación modernos suelen dar servicio a una gran cantidad de usuarios. En esta sección se analizan los rasgos arquitectónicos que se necesitan para facilitar este tipo de operación.

Un **sistema operativo** es un conjunto de programas que hace posible que el usuario de un sistema de computación tenga acceso controlado a sus recursos. Entre los recursos se cuentan la CPU, la memoria y los dispositivos I/O. El sistema operativo asigna tiempo a la CPU, distribuye el espacio de la memoria y asigna y controla dispositivos I/O para cada usuario. Estas funciones se realizan en forma transparente; es decir, el programador escribe el programa como si todo el sistema de computación estuviese dedicado a su trabajo. Para ofrecer características de sistema operativo, la arquitectura de la máquina debe poseer ciertas propie-

dades. Para comenzar, la máquina debe tener cuando menos dos<sup>19</sup> modos de operación diferentes. Uno se denomina modo del supervisor (privilegiado) y el otro modo del usuario (problema).

Cuando la máquina está en modo del supervisor, la CPU puede ejecutar todas las instrucciones de la máquina. Éste es el modo de operación en el cual corren los programas de sistemas operativos, lo que les da un control total del sistema de computación. Todas las peticiones de servicio de los dispositivos periféricos pasan por el sistema operativo, ya que las instrucciones de I/O solamente suelen emitirse cuando la máquina se encuentra en modo del supervisor.

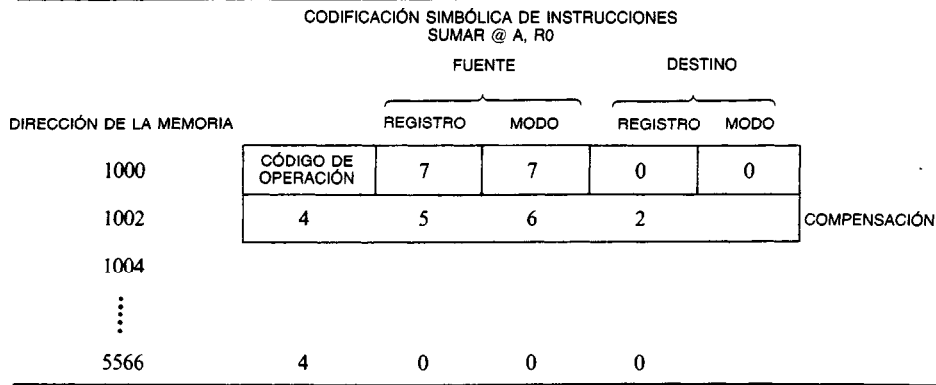
En el modo de problema, la CPU no puede ejecutar todo el conjunto de instrucciones. En particular, no se permite la ejecución de las instrucciones de control de la máquina y de I/O. La restricción impuesta sobre las instrucciones de I/O se hace evidente con un ejem-

DIRECCIÓN DEL PROGRAMA	FUENTE		DESTINO	
	CÓDIGO DE OPERACIÓN	REGISTRO	REGISTRO	MODO
1000		1	7	6
1002	4	5	2	
1004				

COMPENSACIÓN

Fig. 56-19. Direccionamiento relativo, PDP-11. Durante la ejecución de instrucciones, el PC contiene 1 004 direcciones destino efectivas = 1 004 + 4 562 = 5 566 (los valores están en hexadecimal).



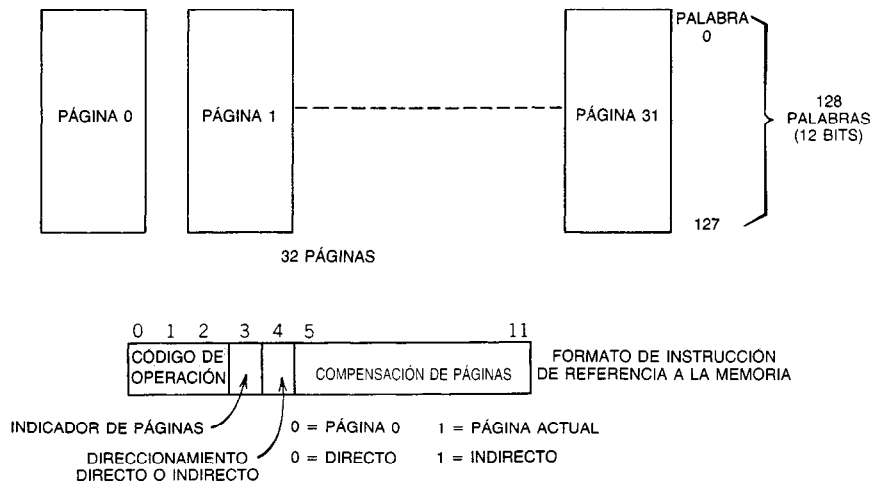


**Fig. 56-20.** Direccionamiento indirecto, PDP-11. Direcciones fuente efectivas = 1 004 + 4 562 = 5 566, pero como se especifica el direccionamiento indirecto (modo de fuente = 7), el operando se encuentra en la localidad 4 000.

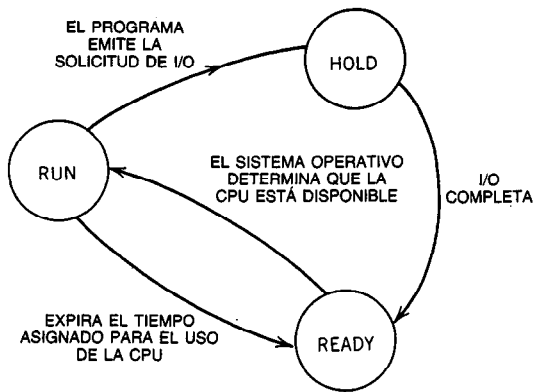
plo. Supóngase que existe una impresora para dar servicio a todos los usuarios del sistema de computación y que la CPU se comparte sobre una base de torneo de eliminación<sup>2,20,21</sup>. Si un usuario imprimiera una línea de salida en su fracción de tiempo asignada y después el usuario siguiente imprimiera también una línea de salida en su fracción de tiempo asignada, la impresión de salida sería un conjunto de líneas entremezcladas de los usuarios del sistema. Es evidente que sería difícil separar las líneas de un usuario de las de otro. Para resolver este problema, cuando un programador requiere I/O, éste debe enviar una instrucción especial al programa de sistemas operativos. Ésta se denomina instrucción de llamada<sup>19</sup> al supervisor (trampa)<sup>4</sup>. El efecto de la instrucción dentro de la CPU consiste en generar una señal interna en el siguiente ciclo de la máquina que haga que ésta cambie su modo por el modo del supervisor. Además, la señal hace que uno de los programas de sistemas operativos dé inicio a la

ejecución. Debido a que la ejecución de la instrucción especial hace que se interrumpa el programa, la señal se conoce como señal de interrupción.

Cuando ocurre la interrupción, el programa de sistemas operativos procede a activar el mecanismo de transferencia de datos con el programa problema. Esto implica asignar espacio de almacenamiento intermedio en la memoria para los datos, suspender la operación del programa interrumpido hasta que se complete la I/O, y asignar la CPU a otro programa. Una herramienta útil para describir el avance de un programa problema en un entorno de usuarios múltiples es un diagrama de estados<sup>19,22,23</sup> (véase la figura 56-22). En el estado RUN (ejecución), se ejecuta el programa; en el estado HOLD (retención), el programa está en espera de que se complete la I/O; en el estado READY (listo), el programa está listo para utilizar la CPU. Los eventos que generan las transiciones de estado también se muestran en la figura.

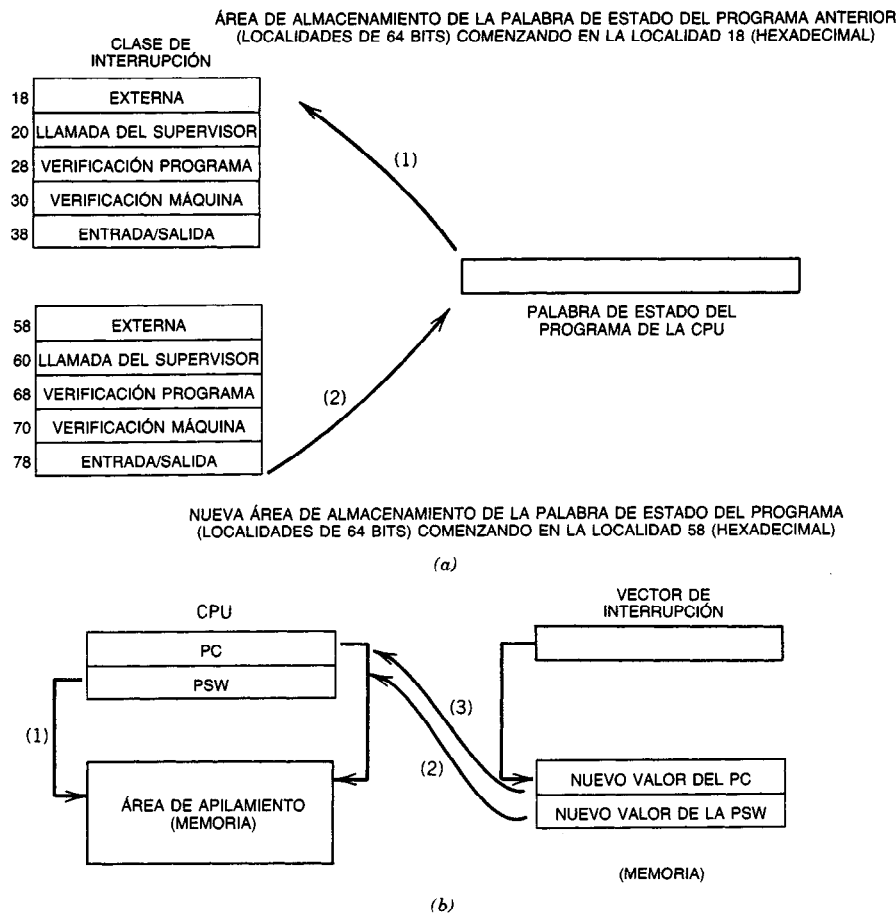


**Fig. 56-21.** Esquema de la memoria, PDP-8.

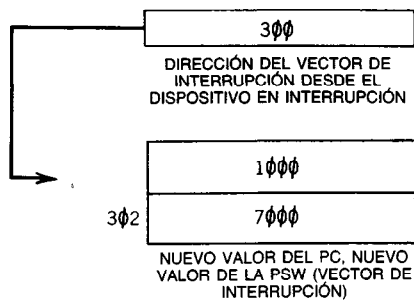


**Fig. 56-22.** Diagrama de estados de un programa en un entorno de compartimiento de recursos. CPU = unidad de procesamiento central; I/O = entrada-salida.

Como se ilustra en el diagrama de estados, cuando se completa la I/O el trabajo avanza al estado READY, y cuando queda disponible la CPU el trabajo vuelve a la ejecución. Para dar inicio al programa problema en su dirección correcta, el sistema operativo deberá retener esa información en el momento en que ocurra la interrupción. La dirección está disponible en el PC en el momento de la interrupción, ya que el PC contiene la dirección de la siguiente instrucción que se ejecutará. La situación es completamente análoga a una rama de subrutina. En el caso de la interrupción, la instrucción de llamada al supervisor transfiere el control a una dirección cableada. En máquinas<sup>3,6</sup> que aceptan instrucciones por pila, el contenido del PC normalmente está almacenado en la pila cuando ocurre la interrupción. En máquinas que no manejan pilas directamente, la dirección de regreso se almacena en un conjunto fijo de localidades de la memoria. En la figura 56-23 se ilustran las acciones de interrupción de



**Fig. 56-23.** Acción de interrupción: a) IBM 370. En la interrupción, la palabra de estado del programa (PSW) de la unidad de procesamiento central (CPU) se almacena en el área de la PSW anterior (1), y la nueva PSW la sustituye (2). Se asigna una palabra de cada área a las cinco interrupciones distintas manejadas por el sistema 370; b) PDP-11. En la interrupción, el contador del programa y la PSW de la CPU se colocan en el área de la pila (1,2). Los nuevos valores se cargan a partir de localidades consecutivas de la memoria, cuya dirección es determinada por la dirección del vector de interrupción. La dirección es devuelta por el dispositivo a la CPU (3) como parte del reconocimiento de la interrupción. PC = contador del programa; PSW = palabra de estado del programa.



**Fig. 56-24.** Interrupción por vectores, PDP-11. La rutina del servicio de interrupciones comienza en la localidad 1000. PC = contador del programa; PSW = palabra de estado del programa.

una PDP-11 (pila) y un sistema 370 (sin pila). Además de la dirección de regreso, también debe almacenarse el estado de la máquina cuando ocurra una interrupción. El estado de la máquina es el contenido de todos los registros de la CPU, así como los códigos de condición cuando ocurre la interrupción. En algunos procesadores<sup>11</sup>, todos los registros se guardan en forma automática, mientras que en otros, el estado de la máquina debe ser guardado por el software. Los dispositivos I/O utilizan asimismo el mecanismo de interrupción para comunicarse con la CPU. La única diferencia entre una interrupción periférica y una interna es que la segunda es sincrónica con el reloj del procesador, mientras que la primera ocurre cuando el dispositivo está listo para aceptar o transmitir información.

Desde el punto de vista del sistema, la mecánica de ambos es la misma. Sin embargo, cuando un periférico genera la interrupción, la dirección inicial del programa de servicio para manejar la interrupción se puede generar en forma externa.

Se aplican dos métodos básicos para identificar la fuente de la interrupción:

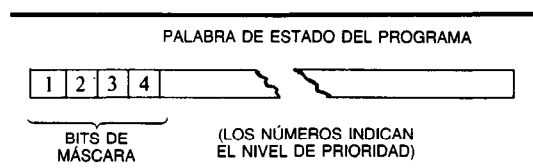
1. En el **método del cableado**, el diseñador del sistema periférico dispone de un conjunto de líneas de señal de interrupción que utiliza como líneas de solicitud de interrupción para el procesador. Cada línea tiene una dirección de memoria particular; asociada a ésta se transfiere el control al recibirse una señal en la línea. El programa de interrupciones debe colocarse en la dirección si el dispositivo emplea la línea.
2. En el **método de los vectores**, el dispositivo periférico proporciona la localidad de la dirección del programa de servicio de interrupciones y también un nuevo valor de la palabra de estado del programa (PSW). En la PDP-11, por ejemplo, estas entidades están cableadas en un par de direcciones asignadas al periférico. Cuando se recibe una solicitud de interrupción en la CPU, ésta lee del dispositivo la dirección del programa de servicio de interrupciones. Se supone que la nueva PSW está ubicada en la siguiente palabra de la memoria. En la figura 56-24 se ilustra el procedimiento.

En un sistema de usuarios múltiples existen muchas fuentes de interrupciones. Por ejemplo, existe un reloj de tiempo real que genera interrupciones en forma periódica, las cuales hacen posible que el sistema operativo realice una planificación. Existen muchos dispositivos I/O, que van desde los de alta velocidad (discos, tambores y cintas) hasta los de baja velocidad (terminales e impresoras). Además, puede haber circuitos que detecten fallas graves del sistema, tales como las de energía. Dado el número de dispositivos, el sistema operativo deberá tener la capacidad de jerarquizar las fuentes de las interrupciones y también la capacidad de ignorarlas.

La estructura de prioridades que se impone a las diversas fuentes depende del uso del sistema. En un sistema de control de tiempo real, un convertidor de analógico a digital (A/D) puede tener prioridad sobre todas las otras interrupciones, de manera que no se pierda ningún dato. En un entorno de un centro de computación, el reloj de tiempo real tendría prioridad sobre la mayor parte de las otras interrupciones, a fin de maximizar el rendimiento del sistema. Naturalmente, una falla de energía u otro tipo catastrófico de interrupción adquiere mayor importancia que las otras.

En muchas máquinas la estructura de prioridades se integra en el hardware; si dos o más interrupciones ocurren al mismo tiempo, se dará servicio a la que tenga la más alta prioridad. Las interrupciones restantes se diferirán. Una técnica común que se aplica para implementar la postergación de interrupciones se ilustra en la figura 56-25. Parte de la PSW funciona como un conjunto de bits "de máscara". Existe un bit para cada nivel de prioridad. Si el bit de máscara es un 1, las peticiones de interrupción de ese nivel se pasan a la CPU, mientras que si el bit es un 0, se pasan por alto las peticiones. Cuando ocurre una interrupción en un nivel determinado, el nuevo valor de la PSW se carga como parte de la secuencia de reconocimiento. Los bits de máscara de la nueva PSW suelen colocarse en 0 en el caso de solicitudes de interrupción de prioridad inferior (o igual) y en 1 en el caso de dispositivos de más alta prioridad.

Si más de un dispositivo comparte las mismas líneas de interrupción en el procesador, se puede construir hardware externo para arbitrar las solicitudes en caso de haber entradas simultáneas. Una alternativa consiste en utilizar un método de interrogación por software. En la figura 56-26 se presenta un programa en lenguaje ensamblador del sistema 370 en el que se emplea interrogación. El programa se introduce cuando el procesador



**Fig. 56-25.** Enmascaramiento de interrupciones. Cuando el bit de máscara es un 1, la interrupción en ese nivel es reconocida por el procesador; cuando es 0, la solicitud se difiere.

DICCIONARIO DE SIMBOLOS EXTERNOS

SYMBOL TYPE ID ADDR LENGTH LD ID  
LIST SD 01 001000 000099

LOC	OBJECT	CODE	ADDR1	ADDR2	STMT	SOURCE STATEMENT
001000					1	LIST START X'1000'
001000	05F0				2	BALR 15,0
001002					3	USING *,15
001002	8000	F096	01098		4	SSM = X'00'
					5	*
001006	4110	F036		01038	6	LA I, READ
00100A	5010	0048		00048	7	ST I, CAW
00100E	9C00	000C	0000C		8	SIO X'00C'
001012	9D00	000C	0000C		9	TIO X'00C'
001016	4770	F010		01012	10	BNZ *-4
					11	*
00101A	4110	F03E		01040	12	LA I, PRINT
00101E	5010	0048		00048	13	ST I, CAW
001022	9C00	000E	0000E		14	SIO X'00E'
001026	9D00	000E	0000E		15	TIO X'00E'
00102A	4770	F024		01026	16	BNZ *-4
00102E	47F0	F004		01006	17	B LOOP
					18	*
001032	000000000000					
001038	0200104800000050				19	READ CCW X'02', BUFFER, X'00', 80 LEER UNA TARJETA
001040	0900104800000050				20	PRINT CCW X'09', BUFFER, X'00', 80 IMPRIMIR UNA TARJETA
001048					21	BUFFER DS CL80
					22	*
000048					23	CAW EQU 72
					24	END
001098	00				25	= X'00'

DICCIONARIO DE RECOLOCACIÓN

POS.ID	REL.ID	FLAGS	ADDRESS
01	01	08	001039
01	01	08	001041

REFERENCIA CRUZADA

SYMBOL	LEN	VALUE	DEFN	REFERENCES
BUFFER	00080	001048	00021	0019 0020
CAW	00001	000048	00023	0007 0013
LIST	00001	001000	00001	
LOOP	00004	001006	00006	0017
PRINT	00008	001040	00020	0012
READ	00008	001038	00019	0006

Fig. 56-26. Interrogación por software, IBM 370. Tomado de S. Madnick y J. Donovan<sup>19</sup>, reimpresso con permiso.

I/O genera una interrupción. Se supone que el procesador de I/O tiene un dispositivo lector de tarjetas y un impresor de líneas conectado a él. Cuando uno de estos dispositivos completa la I/O, genera una interrupción de final del dispositivo. El código de esta interrupción se deja en la palabra de estado del canal<sup>6,6</sup> (CSW). Si la interrupción fue un final de dispositivo, el programa de servicio determina la fuente examinando el código de interrupción, que se encuentra en la PSW de I/O. Después, el programa de servicio reactiva el dispositivo que ha terminado escribiendo la dirección del programa de control de I/O en el área de memoria intermedia de I/O. En las referencias bibliográficas 1 a 4, 11, 13, 16, 18, 19, 24 y 25 se encuentran más detalles sobre la programación de I/O en el sistema 370 y otras CPU.

## 56.6 DISEÑO DE LA UNIDAD DE CONTROL

Para la operación<sup>5,26</sup> adecuada de la máquina se requiere que la transferencia y el procesamiento de información se realicen en una secuencia de tiempo determinada. La función de secuenciación recae en la unidad de control (CU) de la CPU. En esta sección se describen las técnicas de diseño que se aplican a la CU.

En la figura 56-27 se ilustra un diagrama de bloques típico de la CPU, incluyendo los registros y ductos de interconexión. Como las máquinas reales varían mucho en los detalles arquitectónicos, esto solamente se aplica en términos muy generales a una máquina real. Sin embargo, la estructura que se ilustra basta para el análisis de los principios generales de diseño de la CU.

El diagrama muestra un archivo de registros, cuya salida está conectada a las entradas de dos multicanalizadores (LMUX y RMUX). Las salidas de éstos constituyen las entradas a una ALU de funciones múltiples, donde las terminales de entrada FSEL (selección de la función) determinan la función. La salida de la ALU pasa por un desplazador, que puede realizar un desplazamiento de un bit a la izquierda o a la derecha o bien ningún desplazamiento en absoluto. La salida del desplazador se envía a circuitos de desmulticanalización (DEMUX) de manera que se puedan cargar en uno cualquiera de los registros. Las terminales DSEL (selección del destino) seleccionan el registro.

El archivo de registros contiene los MAR, MDR y PC, mientras que el IR es un registro aparte. Las salidas del IR están conectadas a las entradas de la CU.

Además de las salidas del IR, entre las entradas de la CU se cuentan los bits de indicación, líneas de interrupción, y un reloj. Las salidas de la CU controlan el flujo de datos en el registro (sección de la ALU) y entre la CPU y la memoria.

Para ilustrar la función de la CU, considérense las operaciones que se requieren para implementar un ciclo de búsqueda. Supóngase que el PC está contenido en el archivo de registros. La primera tarea por realizar es la transferencia del contenido del PC al MAR. Esto se realiza de la manera siguiente:

1. Se selecciona el PC como una de las entradas a la ALU.

2. Se selecciona un registro que contenga ceros como la otra entrada de la ALU.
3. La operación de la ALU que se elige (vía FSEL) es la adición (de las dos entradas de la ALU). La operación del desplazador seleccionada es la de desplazamiento nulo.
4. La salida del desplazador se dirige a las líneas del MAR (vía el DSEL).

La CU genera las señales de selección y control adecuadas que cierran las compuertas de las trayectorias de datos en forma simultánea, de manera que se efectúe la operación. El paso siguiente consiste en enviar el contenido del MAR al ducto de direcciones y generar una señal de lectura para la memoria. Después se incrementa el PC. Estas operaciones se realizan por medio de un conjunto de subetapas análogas a la transferencia del contenido del PC al MAR.

Cuando se lee el código de operación, éste se transfiere directamente al MDR y después al IR, donde se convierte en una de las entradas a la CU. La unidad de control decodifica la instrucción y ordena las secciones de registro de la ALU para ejecutar la instrucción.

En el ciclo de búsqueda que se describe, se deben ejecutar cuatro pasos en secuencia. Suponiendo que cada uno de los pasos requiere un pulso de reloj para ejecutarse, toda la secuencia requiere cuatro relojes. Dentro de cada paso debe ocurrir otra subdivisión del tiempo. Considérese la transferencia de datos que se realiza entre el PC y el MAR. Para efectuar la operación en forma óptima, primero se deberán formar las líneas de selección de los diversos elementos. Después se puede dirigir la salida de la ALU hacia el MAR. El punto importante es que la selección y transferencia de información no se pueden realizar en forma simultánea, ya que la ALU requiere tiempo para efectuar la adición. Por tanto, la operación requiere dos subciclos (fig. 56-27).

El propio funcionamiento de las compuertas puede ocasionar todavía más problemas. Considérese la actualización del PC. En esta operación, el PC es el registro destino y también el registro fuente. Si la ALU no se diseña con precaución, su salida puede cambiar dos o más veces en el subciclo, provocando lecturas equivocadas del PC. El diseño de la unidad de control requiere un cuidadoso análisis de las sincronizaciones para evitar que se presenten condiciones de competencia, y suele requerir salidas de reloj de fases múltiples junto con las señales de control de las compuertas.

### 56.6.1 Diseño de la CU por cableado

En el método por cableado, la secuenciación se implementa mediante una red de compuertas de control similar a la de la figura 56-27. Cada señal de control se forma aplicando la operación AND\* a una salida del reloj de fases múltiples y una o más señales de control que no son de reloj. En la figura 56-28 se presenta una posible configuración para la generación de las señales de sincronización.

\* El símbolo “.” representa AND.

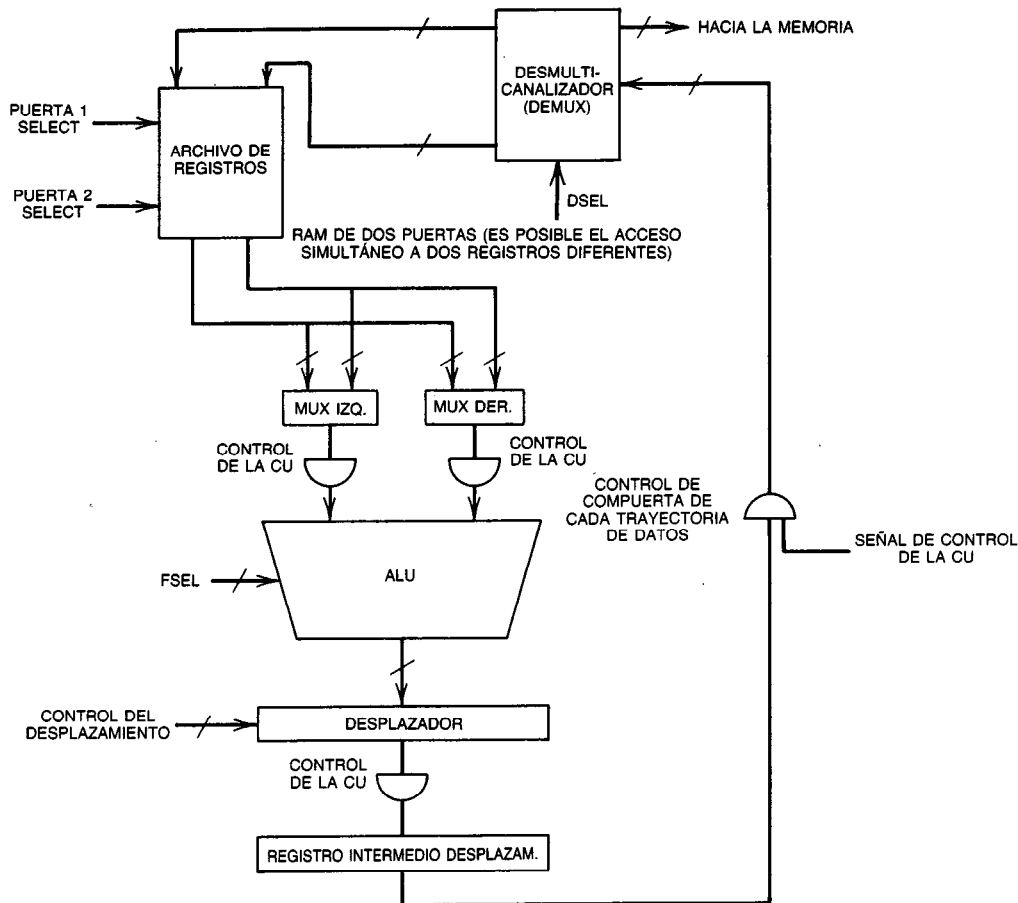


Fig. 56-27. Diagrama de la unidad de procesamiento central (CPU). CU = unidad de control; ALU = unidad de aritmética y lógica; DSEL = selección del destino; FSEL = selección de la función.

El contador anular (o en anillo) de dos estados se acciona a la frecuencia del reloj. El contador de división entre dos acciona un contador anular de cuatro estados. Las salidas del contador de dos estados son las señales de reloj de fases múltiples, y las del contador anular de cuatro estados son las señales de estado de sincronización. Supóngase que estas últimas señales se denominan  $T_0$ ,  $T_1$ ,  $T_2$  y  $T_3$ , y que las subseñales de reloj son  $t_0$  y  $t_1$ . Por tanto,  $T_0 \cdot t_1$  representa el segundo estado de la sincronización en el ciclo del reloj maestro  $T_0$ . Ahora supóngase además que un solo multivibrador biestable distingue entre los ciclos de búsqueda y los de ejecución de la CPU. Por tanto,  $F \cdot T_0 \cdot t_0$  representa el primer subciclo del primer estado de la sincronización dentro del ciclo de búsqueda. Esta señal de control se utiliza para realizar la siguiente operación de compuertas (fig. 56-27).

1. Dirige el PC a las entradas de la derecha de la ALU.
2. Dirige la información de selección a la ALU y al desplazador.

3. Dirige el 0 a las entradas de la izquierda de la ALU.

En el flanco descendente de la subseñal de reloj  $t_0$ , las salidas del desplazador se dirigen al SBR (registro intermedio de desplazamiento).

Durante la subseñal de reloj  $t_1$  de  $T_0$ , el MAR se selecciona en las terminales DSEL, y en el flanco ascendente de la subseñal de reloj  $t_1$  la información se somete a selección estroboscópica en el registro.

Durante el segundo estado de la sincronización ( $T_1$ ), se actualiza el PC y la información de direcciones se coloca en el ducto correspondiente para realizar la búsqueda de instrucciones. Por tanto, se vuelve a seleccionar el PC para su envío a la ALU por medio de la señal de control  $F \cdot T_0 \cdot t_0$ . La trayectoria es cerrada del PC a la entrada de la ALU por  $F \cdot T_0 \cdot t_0$  o por  $F \cdot T_1 \cdot t_0$ ; es decir:

$$ALU_r \leftarrow PC \quad F \cdot t_0 \cdot (T_0 + *T_1)$$

\* El símbolo "+" representa OR lógico.

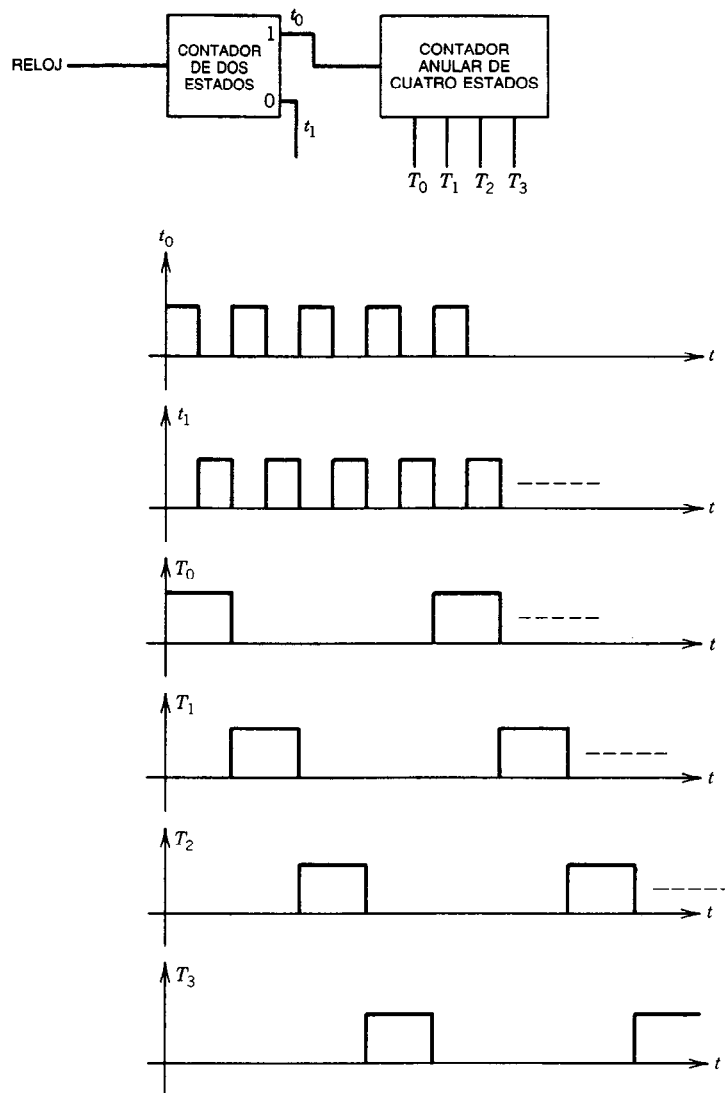


Fig. 56-28. Generación de señales de sincronización para la operación secuencial de la unidad de control.

La mitad izquierda del renglón está en la forma de una instrucción de transferencia<sup>5.22.27</sup> de registros. La mitad derecha del renglón es la señal de control que hace que se habilite la transferencia.

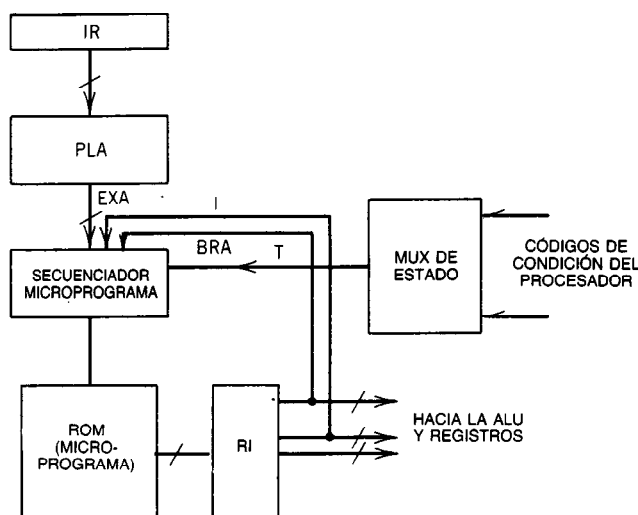
Para completar el diseño de la CU, se enumera cada transferencia de registros necesaria para la ejecución de todas las instrucciones, junto con las señales de control. En una transferencia de registros particular, la señal de control de la trayectoria es generada por el OR de las señales de control asociadas.

Debido al tamaño del conjunto de instrucciones, cada señal de control de la trayectoria constará de muchos términos. Las expresiones lógicas resultantes se implementan mejor mediante el uso de un arreglo de lógica programable (PLA).

### 56.6.2 Diseño de la CU por microprogramación

En esta técnica de diseño, el conjunto de señales de control que se requiere para efectuar la secuencia de operaciones de transferencia de registros (microoperaciones) se almacena en una memoria sólo de lectura (ROM).

En la figura 56-29 se presenta el diagrama de bloques de una unidad de control microprogramada. Los registros de control de la CPU se muestran como un bloque aparte. La salida del IR en este diagrama se aplica a un PLA. Las salidas del PLA se introducen al secuenciador del microprograma. Estas salidas representan la dirección inicial del microprograma asociado con el código de operación del IR.



**Fig. 56-29.** Unidad de control microprogramada. IR = registro de instrucciones; PLA = arreglo de lógica programable; EXA, I, BRA y T = entradas desde diversos registros; ROM = memoria sólo de lectura; MUX = multicanalizador; ALU = unidad de aritmética y lógica; RI = registro intermedio.

El secuenciador del microprograma se conecta a las terminales de dirección de una ROM de control. Esta ROM contiene las microinstrucciones, las señales de control que permiten la ejecución correcta del conjunto de macroinstrucciones de la máquina. Las salidas de datos de la ROM se almacenan en un registro.

Además de las salidas del PLA, el secuenciador tiene otras tres entradas:

- Entradas I, que provienen del campo I del registro intermedio.
- Entradas BRA, del campo de bifurcación de este mismo registro.
- Entrada T (de prueba), que viene del multicanalizador de estado. Las entradas a este multicanalizador son las indicaciones de condición del procesador.

La operación global de la unidad de control es como sigue.

Se busca una instrucción en la memoria y su código de operación se coloca en el IR. El PLA determina la dirección inicial del microprograma para efectuar la operación. Las entradas EXA del secuenciador se almacenan temporalmente de modo interno en un registro denominado contador del microprograma ( $\mu$ PC). Las salidas del  $\mu$ PC se conectan al ducto de direcciones de la ROM.

Después se ejecuta el microprograma en forma secuencial y termina con una instrucción de transferencia que hace que se ejecute el microcódigo del ciclo de búsqueda.

Cada microinstrucción se compone de un conjunto de campos. Los campos se utilizan para funciones definitivas con respecto a la arquitectura controlada. Este aspecto se ilustra regresando a la figura 56-27. Se ne-

cesita un campo para controlar el flujo de datos hacia cada lado de la ALU. El tamaño del campo es determinado por el número de registros que se conmutarán a la ALU. Por ejemplo, si se requiere la conmutación de ocho registros, se necesitan tres bits para cada campo.

Además de los campos para realizar la selección de la ALU, se necesita un campo para la selección del destino y también para la función del desplazador. Además, las funciones de control de la memoria también se controlan por medio de campos.

Cuando se lee la microoperación en el registro intermedio, las salidas seleccionan las funciones adecuadas que se realizarán dentro del ciclo de reloj. El flanco descendente de la señal de reloj provoca la ejecución de las operaciones.

Como ejemplo de microprogramación, considérese nuevamente el ciclo de búsqueda de la CPU. En la tabla 56-2 se presenta una codificación simbólica del microprograma. Los comentarios se explican por sí mismos.

Las operaciones complejas requieren un microcódigo extenso. Al diseñar el algoritmo para la ejecución de una instrucción, suele ser conveniente contar con un medio de ramificación dentro del control microprogramado. Por ejemplo, la última instrucción del ciclo de ejecución de una máquina microprogramada debe ser una ramificación incondicional al microcódigo para el ciclo de búsqueda.

Además, la ramificación condicional puede ser de ayuda en el diseño de algoritmos complejos. Dicha capacidad se implementa en la CPU modelo por medio de una combinación del multicanalizador de estado, los bits de selección del estado (un campo de la microinstrucción) y el campo de ramificación de la microinstrucción.



Tabla 56-2. Ejemplo de microprograma

Formato de la instrucción								
Campos de control (cada uno corresponde a una función de la unidad de control del microprograma [fig. 56-29]).								
	LMUX <sup>a</sup>	RMUX	LSEL	RSEL	SHIFTER	FSEL <sup>b</sup>	DSEL	Comentario
1.	Puerto 1	Puerto 2	PC	0	0	L + R	MAR	MAR ← PC
2.	Puerto 1	Puerto 2	PC	1	0	L + R	PC	PC ← PC + 1
3.	Puerto 1	Puerto 2	MAR	0	0	L + R	Memoria	Memoria ← MAR

<sup>a</sup> LMUX, multicanalizador izquierdo; RMUX, MUX derecho; LSEL, selección izquierda; RSEL, selección derecha; FSEL, función SEL; DSEL, SEL de destino; MAR, registro de direcciones de la memoria; PC, contador del programa.

<sup>b</sup> FSEL: codifica la función de la ALU; es decir, L + R significa sumar la entrada izquierda a la entrada derecha.

Internamente, el secuenciador del microprograma está estructurado como se muestra en la figura 56-30. En base al diagrama de bloques, la dirección de la ROM se selecciona de una de tres fuentes: el  $\mu$ PC, una memoria de pila interna o el campo de ramificación de la microinstrucción. La selección de la fuente es función del ducto I y las líneas SEL, dos de los campos de la microinstrucción.

Las líneas SEL conmutan una de las salidas del código de condición (p. ej., la indicación cero) a la entrada del PLA. El PLA se utiliza para determinar la fuente (vía el multicanalizador) de la siguiente dirección de microinstrucción. Las otras entradas del PLA provienen del ducto I.

Las terminales del ducto I se utilizan para determinar el tipo de microinstrucción. Supóngase que existen tres tipos de microinstrucciones:

1. Un tipo de ejecución, donde la microinstrucción siguiente reside en la localidad que sigue en la ROM.
2. Un tipo de ramificación incondicional, donde la siguiente dirección de la microinstrucción está especificada por el campo de ramificación de la microinstrucción actual.
3. Un tipo de ramificación condicional, donde la dirección de ramificación de la microinstrucción actual se usa si se cumple la condición (si no se cumple, el programa procede en secuencia).

Con estos tres tipos, la dirección de la fuente seleccionada como función del tipo de instrucción y la salida del multicanalizador de estado son como se muestran en la tabla 56-3. El ducto I consta de dos bits ( $I_1$  e  $I_0$ ). El código 00 representa una microinstrucción del tipo de ejecución. En este caso la entrada seleccionada a la ROM es el  $\mu$ PC, que contiene la dirección de la siguiente instrucción en secuencia. (Obsérvese que las salidas de direcciones tienen un 1 sumado, vía el circuito de incremento y se devuelven al  $\mu$ PC.) Así, la entrada seleccionada no es función de la entrada T.

El código 01 representa una ramificación incondicional. Independiente de T, la fuente seleccionada es el campo BRA de la microinstrucción actual.

El código 10 representa una ramificación condicional. La condición probada es programada por el campo

SEL de la microinstrucción actual. Si la condición es verdadera ( $T=1$ ), entonces el campo BRA es la salida, mientras que si la condición es falsa, la salida es el  $\mu$ PC.

La pila permite al microcódigo utilizar subrutinas. (Las direcciones de regreso pueden almacenarse en la pila.) Ahora bien, suponiendo que exista un tipo de instrucción de ramificación de subrutina dentro del conjunto de microinstrucciones, las entradas de la pila al multicanalizador de direcciones se vuelven claras. Para una ramificación de subrutina, se selecciona el campo BRA de la microinstrucción y al mismo tiempo el  $\mu$ PC se coloca en la pila. Para un tipo de instrucción de regreso de subrutina, las entradas de la pila se seleccionan como la fuente de la dirección de la ROM (el equivalente de una operación de extracción de la pila).

Para implementar las características de apilamiento es necesario anexar hardware extra a la estructura secuenciadora. Se necesita un registro indicador de pila para llevar el control de la dirección de la cima de la pila actual. También se necesita otra terminal en el ducto I. Las ramificaciones y retornos de subrutina condicionales se implementan de manera similar a las instrucciones de ramificación condicional.

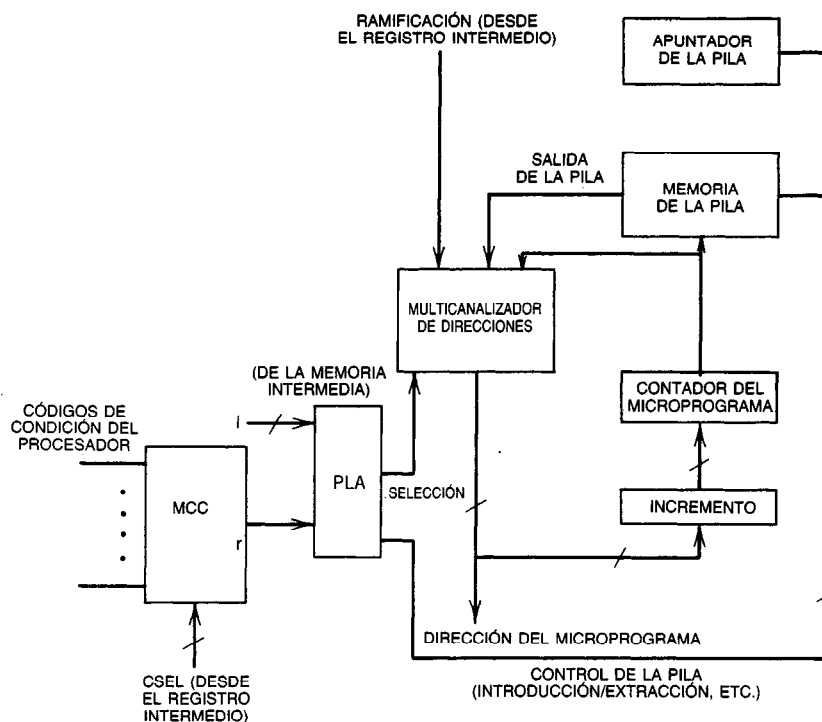
Existen secuenciadores de microprogramas de este tipo en la forma de circuitos integrados a gran escala como una sola unidad. Se fabrican mediante las tecnologías TTL (lógica transistor-transistor) y ECL (lógica acoplada al emisor), para obtener una operación de alta velocidad.

En el capítulo 59 se analiza con mayor profundidad la microprogramación, y las referencias bibliográficas contienen detalles de los circuitos que se utilizan en el diseño de los sistemas<sup>8,9,22,27-30</sup>.

## 56.7 CLASIFICACIÓN DE LOS SISTEMAS DE COMPUTACIÓN

Las arquitecturas de los sistemas de computación han sido clasificadas por Flynn<sup>31</sup>, y esta clasificación se ha utilizado en forma extensiva en la literatura. En esta sección se analiza ese esquema y se dan ejemplos de sistemas de cada una de las clases.

La clasificación se basa en dos ideas: el concepto de flujo de instrucciones y el concepto de flujo de datos.



**Fig. 56-30.** Digrma de bloques de un secuenciador de microprograma. PLA = arreglo de lógica programable; CSEL = selección de la cuenta. MCC = multicanalizador del código de condición.

El primer flujo puede considerarse un programa en ejecución, y el segundo, una serie de datos operados por un programa o conjunto de programas. Los procesadores individuales ejecutan un solo flujo de instrucciones, el lenguaje de máquina del procesador. Los procesadores múltiples pueden ejecutar flujos de instrucciones múltiples, o bien pueden ejecutar el mismo flujo de instrucciones (máquinas idénticas). Los flujos de datos también pueden clasificarse como conjuntos individuales o múltiples de datos.

Al formar todas las combinaciones posibles de los flujos de instrucciones y de datos se obtienen las siguientes clases de sistemas de computación:

1. Una sola instrucción, datos únicos (SISD).
2. Múltiples instrucciones, datos únicos (MISD).
3. Una sola instrucción, datos múltiples (SIMD).
4. Múltiples instrucciones, datos múltiples (MIMD).

Las arquitecturas de uso general que se han descrito hasta ahora entran en las categorías SISD y MISD. La PDP-11, por ejemplo, es una máquina del tipo SISD, ya que todas las instrucciones, incluyendo las de I/O, están contenidas en el conjunto de instrucciones básico del procesador. Para un trabajo particular, los datos necesarios se conservan en un conjunto de datos, de manera consistente con la definición. El procesamiento de un programa en esta máquina es secuencial (en el tiempo).

El sistema 370 de IBM es un ejemplo del tipo de máquina MISD. En este caso el procesador I/O puede considerarse una computadora limitada aparte. Los datos con los cuales operan la CPU y el procesador I/O están contenidos, en relación con un trabajo específico, en un conjunto de datos. El procesamiento de un trabajo en esta máquina también es secuencial.

Los procesadores secuenciales de estos tipos suelen ser adecuados para la implementación de algoritmos secuenciales, y constituyen la mayor parte de los sistemas que están en operación en el presente. En estos sistemas el aumento de la velocidad se debe a mejoras en el hardware más que a cambios estructurales en los algoritmos que ejecutan. Las técnicas básicas de ace-

**Tabla 56-3. Selección de direcciones como función del tipo de instrucción y códigos de condición**

$I_1$	$I_0$	Condición <sup>a</sup>	Dirección de la siguiente microinstrucción <sup>b</sup>
0	0	X	$\mu$ PC (siguiente dirección secuencial)
0	1	X	Entrada BRA (ramificación incondicional)
1	0	1	Entrada BRA
1	0	0	Entrada $\mu$ PC
			} ramificación condicional

<sup>a</sup> Las entradas CSEL seleccionan la condición específica por probar.

<sup>b</sup>  $\mu$ PC, contador del microprograma; BRA, ramificación.

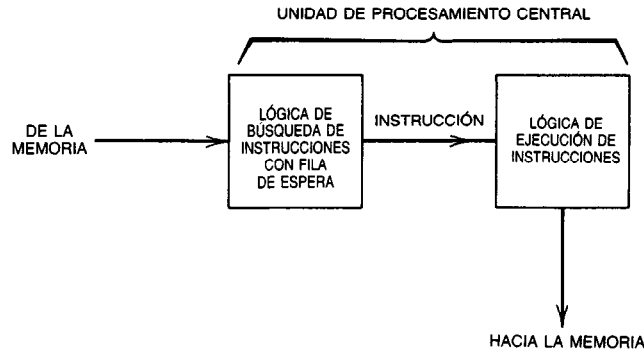


Fig. 56-31. Arquitectura de procesamiento en cadena.

lización del trabajo implican la superposición de los ciclos de búsqueda y ejecución<sup>22</sup> de la máquina y la aplicación de la tecnología de memoria de semiconductores. El método más común de superposición de aquellos ciclos es el procesamiento en cadena. En esta arquitectura, las etapas funcionales implicadas en los dos ciclos de la máquina se asignan a diferentes unidades. El resultado es una línea de ensamblaje secuencial de unidades de hardware, como se muestra en la figura 56-31. Mientras se ejecuta una instrucción, se busca (al mismo tiempo) otra. La concurrencia de operaciones acelera el tiempo promedio de ejecución de un programa. Dos ejemplos de procesadores en cadena son los sistemas 360/195<sup>1</sup> de IBM y el INTEL 8086<sup>11</sup>.

Los sistemas de computación SIMD y MIMD son, desde un punto de vista práctico, sistemas multiprocesadores. La clasificación SIMD<sup>31</sup> relaciona los sistemas multiprocesadores con procesadores idénticos que operan conjuntos de datos físicamente diferentes. Este tipo de procesador también recibe el nombre de procesador de arreglos, y se muestra en forma estructural en la figura 56-32. Esta estructura es más útil para algoritmos que exhiben un alto grado de paralelismo, por ejemplo, problemas de multiplicación de matrices y de predicción de las condiciones climatológicas. Dos ejemplos de este tipo de máquina son el ILLIAC IV<sup>1,32</sup>, producido por la Burroughs Corporation y diseñado en la University of Illinois, y el CRAY-1<sup>12,14,33</sup>.

La clase final de máquina es la MIMD. Si bien éstos son asimismo sistemas multiprocesadores, difieren del tipo MISD en que operan con diversas bases de datos. Un ejemplo importante de esta clasificación es la red de comunicaciones por computadora<sup>1,34-36</sup>. En particular, la red ARPA es aquella en la que diversas computadoras, cada una con bases de datos separadas y sistemas operativos locales, están interconectadas vía canales de comunicaciones. Los canales de comunicaciones constan de líneas telefónicas y radioenlaces (fig. 56-33). Las comunicaciones son manipuladas por minicomputadoras dedicadas a esa tarea. Las computadoras de comunicaciones también se denominan procesadores de enlace de mensajes (IMP). La estructura de la red pone a disposición del suscriptor de la red todo el conjunto de recursos de cada una de las máquinas. Entre otros ejemplos de conexión en redes de

computadoras se cuentan las redes de área local (LAN), el servicio de correo electrónico, los sistemas de apuestas a distancia, y redes electrónicas de finanzas. Las redes de computadoras se analizan en el capítulo 65.

Además de los sistemas geográficamente distribuidos, existen redes de computadoras locales donde las funciones computacionales se distribuyen entre un conjunto de procesadores. Por ejemplo, un sistema de tiempo compartido puede utilizar una microcomputadora para liberar a una máquina anfitriona de la tarea de la comunicación con terminales. La microcomputadora acepta la entrada desde una serie de teclados sobre una base de carácter por carácter, revisa el protocolo y forma mensajes que se enviarán a la máquina anfitriona para su procesamiento.

En el presente la rápida disminución de los costos del hardware debido a la tecnología VLSI (integración a muy grande escala) está haciendo que estos sistemas sean factibles y económicos. Las consideraciones de diseño de los sistemas MIMD incluyen aspectos tales como la estructura de ductos, descomposición funcional y sistema operativo. Se han diseñado y construido algunos sistemas multimicroprocesadores. La referencia 37 contiene una descripción detallada de la implementación del sistema HYDRA, y en la referencia 36 se exploran a fondo los aspectos de diseño que enfrentarán los sistemas futuros.

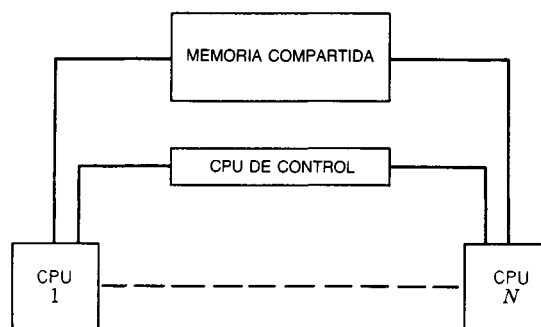


Fig. 56-32. Diagrama funcional de un procesador en paralelo. CPU = unidad de procesamiento central.

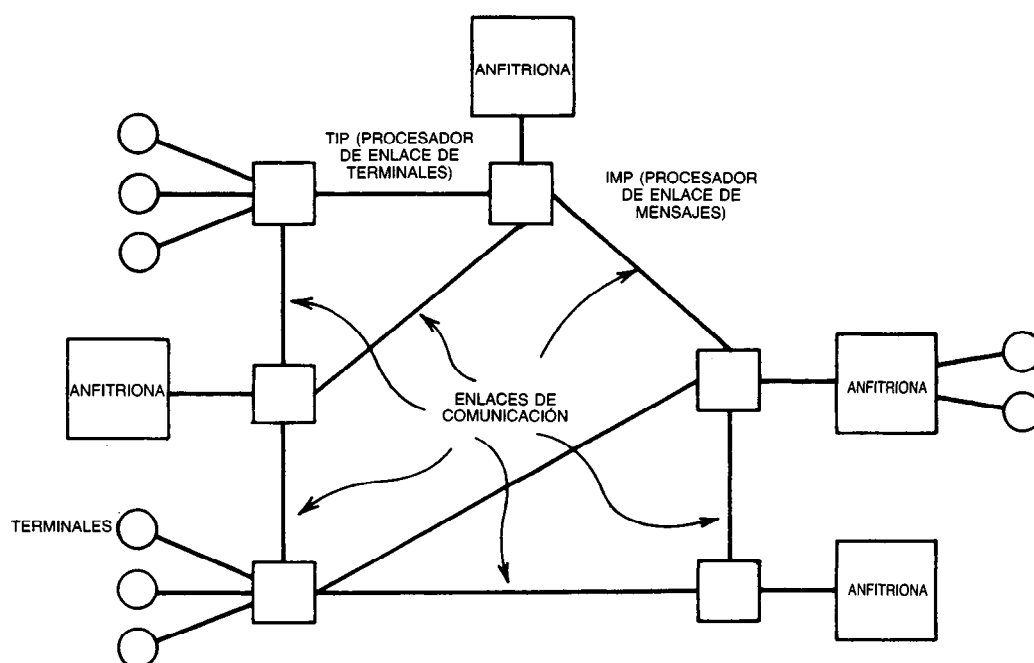


Fig. 56-33. Red de computadoras distribuidas.

Tabla 56-4. Comparación de computadoras

	IBM 370/168	DEC PDP 11/45	Computer Automation NAKED MINI	Intel MCS-80
Costo	\$4.5 millones	\$50 000	\$2 500	\$250
Longitud de palabra (bits)	32	16	16	8
Capacidad de la memoria (bytes de ocho bits)	8.4 millones	256K <sup>a</sup>	64K	64K
Procesador Aumenta tiempo	0.13 $\mu$ s	0.9 $\mu$ s	3.2 $\mu$ s	2.0 $\mu$ s
Velocidad de transmisión de datos de I/O máxima (bytes/segundo)	16 millones	4 millones	1 400 000	500 000
Número de registros de uso general	64	16	3	7
Periféricos (del fabricante)	Todos tipos	Amplia variedad	Disco, cinta, tarjeta impresora de líneas, monitor CRT, casete	Lectora de cinta de papel, disquete, programador de PROM
Software	Todos tipos	Amplia variedad	Sistemas operativos, ensamblador, FORTRAN, BASIC	Ensamblador, monitor, PL/M editor

Fuente: L. Leventhal<sup>38</sup>, reimpreso con permiso.

<sup>a</sup> 1K = 1 024 bits.

## 56.8 COMPARACIÓN DE COMPUTADORAS

En la tabla 56-4 se comparan una computadora grande (IBM 370/168), dos minicomputadoras y un microprocesador. Los parámetros de comparación son los típicos. Con el uso cada vez mayor de los microprocesadores LSI, además de la implementación de máquinas con longitud de palabra grande mediante la aplicación de la lógica de "rebanadas" de bits, la tendencia será hacia una drástica reducción de los costos de las computadoras de alto nivel y una mayor utilización de microprocesadores de bajo nivel en aplicaciones de empresas pequeñas.

El soporte del software y del hardware es creciente en los microprocesadores. Ahora se dispone de varios lenguajes de alto nivel para ellos, como BASIC, FORTRAN, Pascal y COBOL. Además, los sistemas operativos basados en disco son ya muy comunes. Las tecnologías LSI y VLSI harán que los parámetros de comparación se combinen de manera que un sistema independiente resulte indistinguible de otro.

### REFERENCIAS BIBLIOGRÁFICAS

1. A. Tanenbaum, *Structured Computer Organization*, Prentice-Hall, Englewood Cliffs, NJ.
2. C. Bell y A. Newell, *Computer Structures: Readings and Examples*, McGraw-Hill, Nueva York.
3. R. Eckhouse, Jr. y L. Morris, *Minicomputer Systems*, 2.<sup>a</sup> ed., Prentice-Hall, Englewood Cliffs, NJ.
4. Digital Equipment Corporation, *PDP-11/70 Processor Handbook*, Maynard, MA.
5. D. L. Dietmeyer, *Logic Design of Digital Systems*, 2.<sup>a</sup> ed., Allyn & Bacon, Boston, MA.
6. G. Amdahl, G. Blaauw y F. Brooks, Jr., "Architecture of the IBM System/360", *IBM J. Res. Develop.* 8:87-101.
7. W. A. Wulf, M. Shaw, P. N. Hilfinger y L. Flon, *Fundamental Structures of Computer Science*, Addison-Wesley, Reading, MA.
8. J. R. Mick y J. Brick, *Bit-Slice Microprocessor Design*, McGraw-Hill, Nueva York.
9. AMD Corporation, *The 2900 Family Data Book*, Sunnyvale, CA.
10. A. M. Abd-alla y A. C. Meltzer, *Principles of Digital Computer Design*, Vol. 1, Prentice-Hall, Englewood Cliffs, NJ.
11. J. Wakerly, *Microcomputer Architecture and Programming*, Wiley, Nueva York.
12. J. Iliffe, *Advanced Computer Design*, Prentice-Hall International, Londres.
13. Burroughs Corp., *B1700 Reference Manual*, Detroit, MI.
14. R. M. Russell, "The CRAY-1 Computer System", *Comm. ACM* 21(1):63-72.
15. U. W. Pooch y R. Chattergy, *Minicomputers: Hardware, Software and Selection*, West, St. Paul, MN.
16. A. Padegs, "Channel Design Considerations", *IBM Sys. J.* 3(2):165-180.
17. Digital Equipment Corp. *DZ-11 Users Manual*, Maynard, MA.
18. H. S. Stone y D. P. Siewiorek, *Introduction to Computer Organization and Data Structures: PDP-11 Edition*, McGraw-Hill, Nueva York.
19. S. Madnick y J. Donovan, *Operating Systems*, McGraw-Hill, Nueva York.
20. A. C. Shaw, *The Logical Design of Operating Systems*, Prentice-Hall, Englewood Cliffs, NJ.
21. E. G. Coffman, Jr. y P. Denning, *Operating Systems Theory*, Prentice-Hall, Englewood Cliffs, NJ.
22. M. M. Mano, *Digital Logic and Computer Design*, Prentice-Hall, Englewood Cliffs, NJ.
23. V. T. Rhyne, *Fundamentals of Digital Systems Design*, Prentice-Hall, Englewood Cliffs, NJ.
24. G. J. Myers, *Advances in Computer Architecture*, Wiley, Nueva York.
25. R. W. Kline, *Digital Computer Design*, 2.<sup>a</sup> ed., Prentice-Hall, Englewood Cliffs, NJ.
26. J. Donovan, *Systems Programming*, McGraw-Hill, Nueva York.
27. Y. Chu, *Computer Organization and Microprogramming*, Prentice-Hall, Englewood Cliffs, NJ.
28. S. Husson, *Microprogramming: Principles and Practice*, Prentice-Hall, Englewood Cliffs, NJ.
29. M. M. Mano, *Computer System Architecture*, 2.<sup>a</sup> ed., Prentice-Hall, Englewood Cliffs, NJ.
30. F. J. Hill y G. R. Peterson, *Digital Systems: Hardware Organization and Design*, 2.<sup>a</sup> ed., Wiley, Nueva York.
31. M. J. Flynn, "Very High Speed Computing Systems", *Proc. IEEE* 54(12).
32. R. Davis, "The ILLIAC-IV Processing Element," *IEEE Trans. Comp.* C-18(9): 800-816.
33. E. W. Kozdrowicki y D. J. Theis, "Second Generation of Vector Supercomputers", *Computer* 13(11):71-83.
34. P. H. Enslow, Jr., "Multiprocessor Organization—A Survey," *ACM Comp. Sur.* 9:103-129.
35. G. Adams y T. Rolander, "Design Motivation for Multiple Processor Microcomputer Systems", *Comp. Des.* 17:81-89.
36. B. A. Bowen y R. J. A. Brown, *The Logical Design of Multiple-Microprocessor Systems*, Prentice-Hall, Englewood Cliffs, NJ.
37. W. A. Wulf, R. Levin y S. Harbison, *HYDRA/C.mmp An Experimental Computer System*, McGraw-Hill, Nueva York.
38. L. Leventhal, *Introduction to Microprocessors: Software, Hardware, Programming*, Prentice-Hall, Englewood Cliffs, NJ.

## CAPÍTULO 57

# Estructuras de datos

**James A. M. McHugh**

New Jersey Institute of Technology  
Newark, New Jersey

### **57.1 Introducción**

- 57.1.1 Organización secuencial
- 57.1.2 Organización enlazada

### **57.2 Listas lineales enlazadas**

- 57.2.1 Operaciones con listas
- 57.2.2 Listas doblemente enlazadas
- 57.2.3 Listas invertidas
- 57.2.4 Uso apropiado de las listas enlazadas
- 57.2.5 Aplicaciones de las listas enlazadas

### **57.3 Líneas de espera y pilas**

- 57.3.1 Definiciones
- 57.3.2 Líneas de espera
- 57.3.3 Pilas

### **57.4 Árboles**

- 57.4.1 Definición

### **57.5 Tablas de asignación**

- 57.5.1 Definición
- 57.5.2 Uso apropiado
- 57.5.3 Funciones
- 57.5.4 Resolución de colisiones
- 57.5.5 Procedimientos de acceso

## 57.1 INTRODUCCIÓN

Las computadoras procesan información. La estructura y la eficiencia de los algoritmos de procesamiento que se utilizan dependen en alto grado de la forma en que se organiza esta información. Al nivel de la máquina, la información se representa mediante cadenas de bits, bytes y palabras. Al nivel del software, se han generado diversas técnicas de uso general orientadas a los problemas para organizar información, llamadas estructuras de datos. Una **estructura de datos** es una tabla caracterizada por los métodos de acceso y manipulación de su información, su organización de almacenamiento y el tipo de información estructural que contiene.

Los bloques fundamentales de las estructuras de datos suelen ser de tamaños fijos, y se denominan celdas, nodos, átomos, elementos y (en especial para estructuras de datos almacenadas en dispositivos de almacenamiento externos) registros. Las componentes se pueden dividir en partes denominadas campos. Las estructuras de datos se clasifican en dos categorías amplias, secuenciales y enlazadas, según la organización del almacenamiento y sus componentes.

### 57.1.1 Organización secuencial

El arreglo unidimensional se organiza comúnmente como una estructura de datos secuencial. Las componentes del arreglo se almacenan en forma secuencial y contigua. Este orden de almacenamiento facilita el acceso indizado que caracteriza a un arreglo. La  $I$ -ésima componente de un arreglo, donde cada componente consta de  $N$  bytes, se encuentra en una posición fácil de calcular. Suponiendo que  $A$  es la dirección inicial del arreglo, la  $I$ -ésima componente comienza en  $A + N \cdot (I - 1)$ .

### 57.1.2 Organización enlazada

Por otro lado, en las estructuras de datos enlazadas las componentes de la estructura de datos no tienen un orden de almacenamiento específico, y no necesitan ocupar áreas contiguas de la memoria. En este caso la organización es determinada por las direcciones de almacenamiento, o apuntadores, que están contenidas en componentes y que enlazan cada componente con sus componentes estructuralmente asociadas. Por ejemplo, en los árboles, un tipo de estructura enlazada, las componentes forman una jerarquía similar a la de un árbol genealógico, donde las componentes de los padres apuntan a las componentes de los hijos, etc.

En la mayor parte de las estructuras de datos que se describen aquí, la organización enlazada es típica. Algunas, tales como las filas o líneas de espera y las pilas, tienen representaciones secuenciales alternativas. Otras, como las tablas de prueba, tienen representaciones de almacenamiento secuenciales, pero formas de acceso no secuenciales. En lo que sigue se analizan aspectos específicos sobre características y procedimientos de acceso, representación del almacenamiento y áreas de aplicación de las principales estructuras de datos estándares.

## 57.2 LISTAS LINEALES ENLAZADAS

La estructura de datos enlazada más fundamental es la lista lineal enlazada. Cada componente tiene un apuntador único que indica la dirección de una componente sucesora única. Por tanto, la lista es como una cadena. Para localizar una componente, se sigue la pista de apuntadores desde una primera componente hasta la componente objetivo.

La conservación de la estructura de lista conforme se insertan y suprimen componentes de la cadena sólo requiere unos cuantos ajustes de los apuntadores y un mínimo de movimiento de datos. Por otra parte, una lista no permite el tipo de acceso indizado directo que se puede utilizar en un arreglo.

En la figura 57-1 se muestra un diagrama de una lista lineal enlazada. Los recuadros grandes representan las componentes, que constan de uno o más campos de datos (DATA) y un solo campo de apuntador (PTR). La entrada a la lista es vía un apuntador principal (HEAD, cabeza o encabezado) separado, que da la dirección de la primera componente de la lista. Un apuntador nulo (NIL) especial indica el final de la lista. En la figura 57-2 se presenta una implementación mediante el uso de arreglos. Cada componente consta de un renglón de DATA y un renglón correspondiente de PTR (el renglón del apuntador). Los valores del apuntador son números de renglón y no direcciones de almacenamiento absolutas. Un apuntador nulo se representa por medio de 0. El diagrama contiene en realidad dos listas enlazadas: una lista de componentes de repuesto o no usadas con punto de entrada "libre" (FREE), y una lista de componentes en uso o activas con punto de entrada "principal" (HEAD). Si se sigue la pista de apuntadores se ven los renglones 5, 2 y 3 como las componentes activas en el orden de la lista, y los renglones 4 y 1 como las componentes libres en el orden de la lista. Para insertar una nueva componente en la lista activa (punto de entrada principal) se suprime una componente de repuesto de la lista libre (punto de entrada libre), se introducen en su lugar los datos adecuados, y se inserta la componente en la lista activa. El procedimiento inverso se utiliza para suprimir una componente activa.

Este tipo de implemento de arreglo, donde el programador tiene control directo sobre la asignación del almacenamiento, es típico en el lenguaje FORTRAN. Algunos lenguajes de programación, como Pascal y PL/I, tienen características integradas que simplifican la tarea del programador en el manejo del almacenamiento. En la referencia 1 se presenta un análisis de listas y otras estructuras de datos en el contexto de los lenguajes de programación.

### 57.2.1 Operaciones con listas

Las operaciones básicas son recorrido (o traslación), búsqueda, inserción y supresión. Se supone la representación por medio de arreglos de la lista, donde DATA (datos), POINTER (apuntador), HEAD (principal) y FREE (libre) tienen el mismo significado que antes.

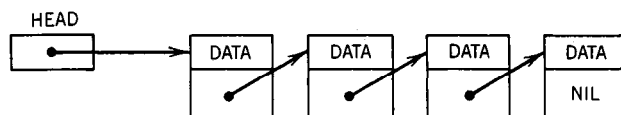


Fig. 57-1. Lista enlazada lineal.

**Recorrido.** El procedimiento de recorrido comienza en el punto de entrada principal de la lista, sigue la pista de apuntadores y procesa componentes que se encuentra en el camino. El pseudocódigo del procedimiento se presenta en la figura 57-3a.

**Búsqueda.** En el procedimiento de búsqueda se hace avanzar el apuntador NEXT (próximo) hasta que se encuentra una componente que coincide con una clave de búsqueda dada, o se regresa un apuntador NIL (nulo) en NEXT si la búsqueda es infructuosa. El pseudocódigo del procedimiento se muestra en la figura 57-3b.

**Inserción.** En este procedimiento es necesario examinar la disponibilidad de una componente libre, localizar el punto de inserción, y efectuar esta última. Si no se dispone de espacio libre cuando se intenta hacer una inserción, se dice que ocurre desbordamiento. Esto suele indicar una condición de error y provoca una salida del procedimiento. El punto de inserción depende del contexto. Por ejemplo, para una lista en orden ascendente de valores de datos, la nueva componente se insertará inmediatamente después de la última componente cuyo valor de datos es menor que el valor de datos de la inserción. Los apuntadores PREV y NEXT se mueven a través de la lista uno tras otro, hasta que PREV apunte a la posición anterior a la inserción y NEXT después de ella. Luego se inserta la nueva componente después de la componente PREV y antes de la NEXT. Los ajustes de apuntadores específicos dependen del punto de la lista en que se realice la inserción. Sea que INSERTKEY denote el valor DATA de la inserción. El pseudocódigo del procedimiento de inserción se muestra en la figura 57-3c.

**Supresión.** En la figura 57-3d se presenta el pseudocódigo del procedimiento que suprime la componente posterior a la señalada por P. La componente suprimida OLD puede insertarse en la lista libre o en otra lista, según convenga.

### 57.2.2 Listas doblemente enlazadas

En una lista doblemente enlazada, cada componente apunta hacia adelante a su sucesora y también hacia atrás a su predecesora, con apuntadores de cabeza y cola separados, que direccionan la primera y última componentes. Esta lista se puede recorrer en una dirección u otra desde una componente cualquiera. Esto simplifica operaciones como la inserción. En una representación estándar los enlaces dobles se implementan con un par de apuntadores en cada componente. En una representación alternativa se combinan las

direcciones del sucesor y del predecesor en un solo campo apuntador, aplicando la operación OR excluyente de sus representaciones de cadenas de bits. Utilizar un solo apuntador representa una reducción en el número total de componentes que se pueden direccionar y un procedimiento de recorrido algo más complejo<sup>2</sup>.

### 57.2.3 Listas invertidas

La búsqueda en una lista es más eficiente si ésta se conserva ordenada por clave de búsqueda. Así, un directorio telefónico de clientes podría ordenarse por el nombre del cliente, lo cual facilita la forma de búsqueda más usual. Pero también podría necesitarse encontrar el nombre de un cliente dado, un domicilio o un número telefónico. Una solución consiste en conservar muchas listas, cada una en orden conforme a una clave diferente. Una solución de almacenamiento eficiente consiste en tener una sola lista y utilizar campos de apuntadores múltiples para codificar los diferentes órdenes de claves. Las componentes podrían tener los campos NAME (nombre), ADDRESS (domicilio), TELEPHONE (teléfono), NAMEPTR, ADDRPTTR y TELEPTR, donde cada campo de apuntador indique un sucesor alfanumérico con respecto a ese campo. NAMEPTR indica la dirección de la componente con el siguiente nombre en orden alfabético, etc. También se requieren apuntadores separados a la cabeza de cada lista. Los procedimientos de conservación de estas listas invertidas (o listas multienlazadas) son similares a los de una lista de enlace sencillo<sup>3,4</sup>.

### 57.2.4 Uso apropiado de las listas enlazadas

El uso de las listas enlazadas es apropiado si se insertan y suprimen componentes con frecuencia, si se espera que las necesidades de almacenamiento y la forma de la tabla de datos varíen mucho durante la ejecución del programa, o si las capacidades de apuntadores de una lista capturan características estructurales importantes de un problema. Por ejemplo, considérese un sistema de reservaciones de una línea aérea que implica la colocación de pasajeros en distintos vuelos. Una

	DATA	PTR	
1	—	0	HEAD
2	12.3	3	5
3	42.5	0	FREE
4	—	1	4
5	83.6	2	

Fig. 57-2. Implementación con arreglos de una lista enlazada lineal.



posible estructura de datos consta de una lista enlazada con componentes de información sobre vuelos separadas para cada vuelo. A cada componente de vuelo están anexadas dos sublistas: una lista de los pasajeros del vuelo y una lista de espera. Las listas crecen y decrecen rápidamente a medida que se crean vuelos, se cancelan reservaciones, los pasajeros pasan de la lista de espera a la de vuelo programado, etc. Los procedimientos de mantenimiento de una organización enlazada hacen posible la inserción y supresión eficientes de componentes con el uso de unos cuantos cambios en los apuntadores. Los requisitos de forma y almacenamiento de la tabla de datos son cambiantes, ya que tanto el número como el tamaño de los vuelos y sublistas varían enormemente. Una vez más, la estructura enlazada resulta adecuada debido a que permite el uso eficiente del almacenamiento. Las localidades de almacenamiento que quedan libres conforme se suprimen componentes de una parte de la estructura pueden ser reutilizadas por componentes recién creadas anexadas a otra parte. Por último, la correspondencia natural entre la estructura de datos y el problema debe simplificar la evolución del programa<sup>5,6</sup>.

Cuando la elección de la estructura de datos está entre un arreglo lineal ordenado con base en alguna clave y una lista lineal enlazada con la misma clave, es posible dar lineamientos más específicos. Básicamente, el arreglo es superior en cuanto a búsqueda y ordenamiento, mientras que la lista es más adecuada para dar cabida al mantenimiento dinámico (es decir, la inserción y supresión de componentes). Por tanto, un arreglo con  $N$  componentes se puede ordenar en  $N \log_2(N)$  operaciones, y una búsqueda binaria localiza una componente con una clave dada a lo sumo en  $\log_2(N)$  pasos<sup>7</sup>. Sin embargo, para la inserción de una nueva componente en un arreglo ordenado es necesario mover un promedio de  $N/2$  componentes a fin de dejar espacio para la nueva componente. El rendimiento en la búsqueda de una lista enlazada es afectado por la naturaleza estrictamente secuencial del acceso a la lista. Para localizar una componente con una clave elegida al azar se requiere el acceso a un promedio de la mitad de las componentes de la lista. Pero la inserción de una componente, una vez que se ha identificado el punto de inserción, puede realizarse con unos cuantos ajustes de los apuntadores. Una regla general es que, en el caso de una población estática conservada en orden con alguna clave y sujeta a búsqueda frecuente pero con algunas inserciones o supresiones esperadas, resulta apropiado un arreglo ordenado. Sin embargo, en un entorno dinámico la lista suele ser superior<sup>8</sup>.

### 57.2.5 Aplicaciones de las listas enlazadas

El manejo de la asignación y desasignación de bloques de almacenamiento de tamaño variable se denomina manejo dinámico de la memoria. Éste es un problema que ocurre con sistemas operativos que solicitan espacio para realizar nuevos trabajos y con programas que construyen estructuras de datos enlazadas. Inicialmente, se comienza con un solo bloque grande de espacio

libre. A medida que crecen las solicitudes de almacenamiento, se separan bloques menores de este espacio libre. Los bloques asignados en un momento se devuelven al espacio libre en tiempos arbitrarios posteriores. Conforme progresan las asignaciones y desasignaciones, el área libre original se vuelve un "tablero de ajedrez" con espacios libres y en uso. Para manejar esta configuración, los bloques de almacenamiento, sean asignados o libres, se toman como componentes de una lista doblemente enlazada. Los bloques figuran en la lista en el orden de su dirección inicial. Se reserva una pequeña cantidad de espacio en cada bloque para apuntadores dirigidos a sus bloques sucesor y predecesor y también para un bit de estado que indica si el bloque está libre o en uso. Para asignar un bloque de  $M$  bytes se requiere buscar en la lista de almacenamiento un bloque libre de tamaño  $N(\geq M)$ . Si no se dispone de este bloque, se rechaza la solicitud. De lo contrario, el bloque localizado se divide en dos bloques de tamaños  $M$  y  $N-M$ . Uno se asigna, el otro permanece libre; los bloques se enlazan y sustituyen la componente o bloque individual del cual se derivaron. Para desasignar un bloque, se entra a sus bloques sucesor y predecesor (que son áreas de almacenamiento limítrofes). Si uno de estos bloques está libre, se combina con el bloque que se está desasignando en un bloque libre de mayor tamaño. Esto simplifica la asignación de espacio subsiguiente<sup>9</sup>.

Algunas políticas de manejo del almacenamiento no devuelven de inmediato bloques desasignados a una lista libre. En cambio, se invoca un procedimiento llamado "de recolección de basura" cuando la cantidad de espacio en la lista libre se vuelve precariamente escasa. El recolector de basura busca en la memoria bloques que estén siendo apuntados actualmente por alguna estructura de datos y los marca como en uso. Los bloques no marcados se recuperan reintegrándolos a una lista de bloques libres<sup>2,8</sup>.

Un área de aplicación que ha constituido simultáneamente un estímulo importante a la teoría de las estructuras de datos es la de las gráficas interactivas<sup>3,10-12</sup>. Las estructuras de datos que se emplean dependen en alto grado del sistema de gráficas específico, pero la idea general es la de representar un despliegue de gráficas por medio de una estructura de datos, donde cada elemento de margen del despliegue esté definido por los parámetros de una componente correspondiente de la estructura de datos. Por ejemplo, suponiendo que las órdenes para el trazado de líneas son instrucciones primitivas del sistema de despliegue, una componente específica podría indicar que debe trazarse una línea de cierta longitud y ángulo desde una posición actual. Un procedimiento de despliegue de software explora la estructura de datos y utiliza el contenido de las componentes para identificar, fijar parámetros de las instrucciones de despliegues de hardware adecuados, y ejecutar dichas órdenes.

Se pueden emplear listas enlazadas para representar de manera eficiente arreglos dispersos y expresiones simbólicas, tales como los polinomios. Un arreglo es disperso si la mayor parte de sus entradas son cero. En una representación con uso eficiente del almace-

namiento de un arreglo disperso emplea una lista enlazada, cada una de cuyas componentes corresponde a una entrada distinta de cero del arreglo y contiene renglón, columna y valor de la entrada, y apuntadores a las siguientes componentes distintas de cero de su renglón y columna. El acceso a componentes individuales es más lento que para la representación secuencial estándar de un arreglo, pero se conserva el espacio y las operaciones del arreglo se pueden implementar de manera eficiente<sup>3</sup>. La manipulación simbólica de polinomios también puede basarse en la representación de una lista. Cada componente de la lista contiene el coeficiente y exponente de un término del polinomio y un apuntador a la componente del término que sigue. Las operaciones con polinomios estándares se implementan en forma eficiente por operaciones de listas típicas<sup>9</sup>.

Otra aplicación de las listas es la edición de texto, en la que el texto se considera una lista enlazada de caracteres (o bloques de caracteres), y operaciones de procesamiento de texto, tales como la inserción, supresión y sustitución de líneas o cadenas de texto, se facilitan con la representación por lista. Ésta y otras aplicaciones se tratan en las referencias 13 a 16.

## 57.3 LÍNEAS DE ESPERA Y PILAS

### 57.3.1 Definiciones

Varios tipos de listas importantes tienen puntos de acceso controlados, y solamente se pueden hacer inserciones y supresiones en puntos específicos. Algunos de estos tipos son:

- **Líneas de espera (filas o colas).** Las inserciones se hacen en un extremo de la lista y las supresiones se hacen en el otro extremo.
- **Pilas.** Las inserciones y supresiones se realizan en un solo extremo de la lista.
- **Deque.** Las inserciones y supresiones pueden realizarse en un extremo u otro de la lista.

Los algoritmos para los deque son simples extensiones de los propios de las líneas de espera, por lo que no se considerarán más a fondo. Knuth<sup>8</sup> presenta mayor información.

### 57.3.2 Líneas de espera

Una línea de espera (también llamada fila o cola) es como una fila en un contador de revisión. Las llegadas se insertan en un extremo, llamado cola de la línea de espera, y en el otro extremo, la cabeza de la línea de espera, se procesan o suprimen componentes.

**Uso apropiado.** Las líneas de espera se utilizan cuando la demanda de un servicio puede superar la capacidad de un sistema de procesar peticiones, cuando las peticiones de servicio excedentes se acumulan en vez de rechazarse, y cuando se debe conservar el orden de llegadas de las peticiones. Situaciones de este tipo son comunes en la programación de sistemas. Por ejemplo,

---

Set NEXT := HEAD.

WHILE NEXT ≠ 0 DO:

    Process DATA(NEXT).

    Set NEXT := PTR(NEXT).

END DO.

(a)

Set NEXT := HEAD and FOUND := FALSE.

WHILE NEXT ≠ 0 AND FOUND = FALSE DO:

    IF DATA (NEXT) ≠ SEARCHKEY

        set NEXT := PTR(NEXT),

    ELSE

        set FOUND := TRUE.

END DO.

(b)

IF FREE ≠ 0 (\* ¿HAY ESPACIO DISPONIBLE? \*)

    Set NEW := FREE and FREE := PTR(FREE).

    (\* ENCONTRAR LA POSICIÓN DE LA INSERCIÓN. \*)

    Set PREV := 0, NEXT := HEAD, and FOUND :

        = FALSE.

    WHILE NEXT ≠ 0 and FOUND = FALSE DO:

        IF DATA(NEXT) < INSERTKEY

            set PREV := NEXT and NEXT := PTR(NEXT).

        ELSE

            set FOUND := TRUE.

    END DO.

    (\* HACER LA INSERCIÓN. \*)

    IF PREV ≠ 0

        set PTR(PREV) := NEW and PTR(NEW) := NEXT,

    ELSE

        set PTR(NEW) := HEAD and HEAD := NEW.

ELSE (\* LISTA LIBRE VACÍA. \*)

    COLOCAR UN MENSAJE DE DESBORDAMIENTO

(c)

IF PTR(P) ≠ 0

    set OLD := PTR(P) and PTR(P) := PTR(OLD).

(d)

---

**Fig. 57-3.** Algoritmos de listas enlazadas: a) recorrido; b) búsqueda; c) inserción; d) supresión. Set = hacer; process = procesar; and = y.

los sistemas operativos mantienen líneas de espera de trabajos que esperan ser programados. También existen líneas de espera de entrada/salida (I/O), en las que los programas colocan las solicitudes de transferencia de datos hacia o desde dispositivos I/O, y líneas de espera de tareas, en las que los programas listos para correr esperan acceso a la CPU.

A menudo resulta apropiado conservar líneas de espera múltiples, cada una correspondiente a un grupo de peticiones acumuladas de prioridad dada. Un procesador manejaría todas las peticiones de una línea de espera de mayor prioridad antes que ninguna otra solicitud de una línea de espera de prioridad inferior. Un conjunto de líneas de espera, ordenadas por prioridad, también pueden implementarse mediante el uso de una estructura de datos llamada montón. Los montones tienen la ventaja de que las inserciones y supresiones se pueden efectuar en  $\log_2(N)$  operaciones, donde  $N$  es el número de entradas en el montón<sup>2</sup>.

Quizá la aplicación clásica de los montones sea el almacenamiento temporal de entradas y salidas. El problema que ocurre es cómo vincular un dispositivo de procesamiento de alta velocidad, tal como una CPU, con dispositivos I/O de menor velocidad, tales como discos o impresoras. En forma específica, un programa que se ejecuta en una CPU puede generar salida para una impresora más aprisa de lo que la impresora puede manipular las peticiones. En consecuencia, los registros de salida deberán almacenarse en forma temporal y, por último, transmitirse en ese orden de llegadas. Ésta es precisamente la circunstancia en la cual una línea de espera resulta apropiada. Típicamente, la línea de espera de registros de salida reside en un área de almacenamiento llamada memoria intermedia, que corresponde en tamaño a la cantidad exacta de información que el dispositivo I/O puede transferir en un momento dado. Un conjunto de estas memorias intermedias puede enlazarse entre sí en una lista circular llamada depósito intermedio. En un momento cualquiera, algunas de estas memorias están llenas y en espera de la transmisión. Otras están vacías y en espera de ser llenadas por la CPU. Mientras tanto, una memoria intermedia parcialmente llena sigue llenándose con la salida. Como la CPU hace adiciones a la cola de esta línea de espera de salida y el dispositivo I/O transmite desde su cabeza, la CPU y el dispositivo I/O pueden competir entre sí en forma efectiva en torno a la lista circular de memorias intermedias I/O. Siempre que se agotan las memorias intermedias vacías de la CPU, debido a que se están produciendo registros de salida más aprisa de lo que el dispositivo I/O los puede procesar, la CPU se bloquea (hasta el momento en que aquel dispositivo libere una memoria intermedia al transmitir sus datos). A la inversa, si el dispositivo I/O alcanza a la CPU, espera una señal de que se ha llenado otra memoria antes de proseguir la transmisión, de manera que se evita transmitir memorias intermedias vacías o parcialmente llenas<sup>3</sup>.

**Representaciones y algoritmos de acceso de líneas de espera.** Las líneas de espera pueden implementarse mediante representaciones enlazadas o secuenciales.

Las primeras tienen la ventaja de que las componentes pueden ser insertadas o suprimidas exclusivamente por cambios de los apuntadores, un mínimo de movimiento de datos. Esto es importante en aplicaciones en que las componentes emigran de una línea de espera a otra. Por otra parte, la implementación secuencial no requiere el mantenimiento explícito de una lista libre de espacio de almacenamiento de respaldo para componentes futuras, y también hace posible un recorrido secuencial más rápido de la línea de espera. Dos representaciones enlazadas son estándares, a saber, la representación lineal y la línea de espera circular.

En la representación enlazada lineal, apuntadores separados de cabeza y cola identifican la primera y última componentes de la línea de espera, y cada componente apunta a su sucesor en el orden de la línea de espera, donde la última componente tiene un apuntador nulo.

En una representación enlazada circular, todo el acceso es a través de un apuntador de cola único. Cada componente apunta a su sucesor en el orden de la línea de espera, salvo la última componente que apunta a la primera componente de la línea. El acceso a la cabeza de la línea de espera es indirecto: primero se entra a la última componente vía TAIL (cola) y después se sigue al apuntador de esa componente para localizar la cabeza. El costo de este acceso indirecto a la cabeza de la línea de espera se compensa con la simplificación resultante en los algoritmos de acceso.

A continuación se presentan los algoritmos de inserción y supresión de la representación enlazada circular de una línea de espera. Los algoritmos de la representación enlazada lineal son similares y, en consecuencia, se omiten.

1. **Inserción (representación enlazada circular).** Sean TAIL, DATA y PTR los campos de cola, datos y apuntador de una línea de espera circular. Sea NEW una componente libre obtenida de una lista libre conservada por separado. Sea VALUE los datos que se insertarán. El pseudocódigo del procedimiento para insertar NEW en la línea de espera se ilustra en la figura 57-4a.
2. **Supresión (representación enlazada circular).** Para suprimir una componente de la (cabeza de la) línea de espera, se utiliza el pseudocódigo de la figura 57-4b.

Las líneas de espera también pueden implementarse mediante el uso de una representación de almacenamiento secuencial, en contraste con las representaciones enlazadas lineales o circulares. Los algoritmos apropiados para esta representación se explican enseguida. Sea  $Q$  un arreglo lineal de tamaño  $N$  e inicialícese HEAD y TAIL a  $N$ .

1. **Inserción (representación secuencial).** El pseudocódigo del procedimiento para insertar un elemento en una línea de espera representada en forma secuencial se ilustra en la figura 57-5a. Ocurre desbordamiento cuando no se dispone de espacio para realizar la inserción.

---

```

Set DATA(NEW) := VALUE.

IF TAIL = NIL                                (* ¿ESTÁ VACÍA LA FILA DE ESPERA? *)

    set TAIL := NEW and PTR(NEW) := NEW.

ELSE

    Set PTR(NEW) := PTR(TAIL).
    set PTR(TAIL) := NEW, and                (* HACER QUE NEW SEA LA NUEVA COLA. *)
    set TAIL := NEW.

(a)

IF TAIL = NIL                                (* ¿ESTÁ VACÍA LA FILA DE ESPERA? *)

    set P := NIL.

ELSE

    Set P := PTR(TAIL).                      (* IDENTIFICAR LA CABEZA. *)

    IF PTR(TAIL) = TAIL                      (* RETIRAR LA CABEZA. *)

        set TAIL := NIL.

    ELSE

        set PTR(TAIL) := PTR(P).

(b)

```

---

**Fig. 57-4.** Algoritmos de líneas (filas) de espera circulares: a) inserción; b) supresión.  
Set = hacer; and = y.

2. **Supresión (representación secuencial).** El pseudocódigo del procedimiento de supresión se presenta en la figura 57-5b. Ocurre insuficiencia cuando se intenta suprimir un elemento de una línea de espera vacía.

Los algoritmos de representación secuencial de la figura 57-5 son breves pero sutiles. TAIL apunta a la última componente, mientras que HEAD apunta a la posición anterior a la primera componente. A lo sumo, se pueden almacenar  $N-1$  componentes a la vez, aunque el arreglo subyacente sea de tamaño  $N$ . El arreglo  $Q$  actúa como una lista circular. INSERT mueve a TAIL en sentido horario (del reloj) y DELETE mueve a HEAD en ese mismo sentido. Ocurre desbordamiento cuando TAIL alcanza a HEAD. Ocurre insuficiencia cuando HEAD alcanza a TAIL. Los desbordamientos suelen estar considerados como condiciones de error que normalmente deberán provocar la terminación del programa, mientras que la insuficiencia suele ser una condición importante que refleja la terminación del procesamiento de la línea de espera.

### 57.3.3 Pilas

Las pilas son listas en las cuales sólo se pueden hacer inserciones y supresiones en un extremo. La operación de inserción se denomina PUSH (introducir o cargar) y

la operación de supresión se conoce como POP (extraer o descargar). El extremo distinguido de la pila se llama cima (TOP). Una pila se comporta en forma muy semejante a una pila de platos ordinaria. Los platos se agregan o retiran fácilmente de la cima de la pila, pero es difícil hacer lo mismo de las partes intermedia o inferior.

**Uso apropiado.** Las pilas se forman siempre que se procesan datos en el sentido contrario al de su orden de generación. Por ejemplo, considérense las llamadas subrutinas. Cuando una subrutina llama a otra, tiene que rescatarse o guardarse información acerca de la primera. Esto incluye la dirección de retorno, o sea el punto en la rutina solicitante al cual deberá regresar el control al término de la rutina llamada. Si ocurre una secuencia anidada de llamadas a subrutinas, donde la rutina S1 llama a la rutina S2, que a su vez llama a la rutina S3, etcétera, tiene que guardarse una secuencia completa de direcciones de retorno (una dirección para cada una de las invocaciones).

Como las subrutinas se devuelven exactamente a la inversa de su orden de petición, una pila representa el mecanismo de almacenamiento adecuado. Los valores de registros generales también suelen apilarse. Esto da a la rutina llamada acceso libre a los registros, cuyos valores originales se pueden reestablecer a partir de la copia apilada al regreso. Por tanto, las llamadas de

```

IF TAIL = N
    set TAIL := 1,
ELSE
    set TAIL := TAIL + 1.
IF TAIL ≠ HEAD
    set Q(TAIL) := VALUE,
ELSE
COLOCAR UNA ETIQUETA DE DESBORDAMIENTO
    (a)
IF TAIL ≠ HEAD
    IF HEAD = N
        set HEAD := 1,
    ELSE
        set HEAD := HEAD + 1.
    Set VALUE := Q(HEAD).
ELSE
COLOCAR UNA ETIQUETA DE INSUFICIENCIA
    (b)

```

Fig. 57-5. Algoritmos de líneas de espera secuenciales: a) inserción; b) supresión. Set = hacer; and = y.

subrutinas requieren la introducción de parámetros en una pila, mientras que el regreso de una subrutina acarrea parámetros de descarga a fin de restablecer el entorno original<sup>17</sup>.

Se puede aplicar la misma idea general para implementar rutinas recurrentes, en donde a una subrutina se le permite llamarse a sí misma<sup>1</sup>.

Las pilas también se utilizan cuando ocurren interrupciones. Una interrupción es una señal para la CPU, generada por un dispositivo o programa, a fin de que opere el servicio del sistema. Las interrupciones provocan una transferencia automática del control de un programa cualquiera que esté en ejecución actualmente a una rutina el sistema operativo. Por tanto, las interrupciones tienen un efecto similar a una llamada de subrutina, y las pilas se utilizan aquí por las mismas razones<sup>18</sup>.

Las pilas son muy utilizadas en la compilación de lenguajes de nivel superior. Una ilustración muy simple de su extenso empleo es la evaluación de expresiones aritméticas escritas en forma posfija, donde los operadores siguen a sus operandos. Por ejemplo, la expresión algebraica ordinaria  $(A+B)*(C-D)$  se escribirá en forma de posfijo como:  $AB+CD-*$ . Para

evaluar esta expresión, se explora de izquierda a derecha y los operandos se introducen en una pila conforme se encuentran. Cuando se explora un operador, se descargan los dos operandos de la cima, se aplica el operador y el resultado se introduce en la pila. Al terminar, la cima de la pila es el resultado de la evaluación de la expresión<sup>17</sup>.

Algunas otras aplicaciones de las pilas son el ordenamiento, la asignación de almacenamiento dinámico en lenguajes estructurados por bloques, como PASCAL, y algoritmos de recorrido de árboles<sup>1</sup>. Horowitz y Sahni<sup>9</sup> dan un ejemplo instructivo del uso de las pilas para resolver un problema de recorrido de laberintos.

**Representaciones y algoritmos de acceso de pilas.** Las pilas, al igual que las líneas de espera, tienen representaciones enlazadas y secuenciales. En el caso de las pilas, la representación secuencial es la más común. Sea  $S$  un arreglo lineal de tamaño  $N$ . Inicialícese TOP a cero. Los pseudocódigos de los procedimientos PUSH y POP se muestran en la figura 57-6.

Algunas aplicaciones requieren un par de pilas complementarias, donde una crece cuando la otra decrece y a la inversa. Estas dos tareas se pueden representar en forma adecuada en un área compartida de almacenamiento, fijando las bases de las pilas en extremos opuestos de la región compartida. Las técnicas de almacenamiento compartido entre tres o más pilas son más sutiles<sup>2</sup>.

## 57.4 ÁRBOLES

### 57.4.1 Definición

Un árbol es una clase de estructura de datos enlazada. En la figura 57-7 se da un ejemplo. La figura representa una organización de datos jerárquica, tal como podría ocurrir en un medio bancario. Las componentes de diferentes niveles de la jerarquía contienen diferentes tipos de información. La componente más alta con-

```

IF TOP < N
    set TOP := TOP + 1 and S(TOP) := VALUE,
ELSE
COLOCAR UNA ETIQUETA DE DESBORDAMIENTO
    (a)
IF TOP > 0
    set VALUE := S(TOP) and TOP := TOP - 1,
ELSE
COLOCAR UNA ETIQUETA DE INSUFICIENCIA
    (b)

```

Fig. 57-6. Algoritmos de pilas: a) introducción; b) descarga. Set = hacer; and = y.

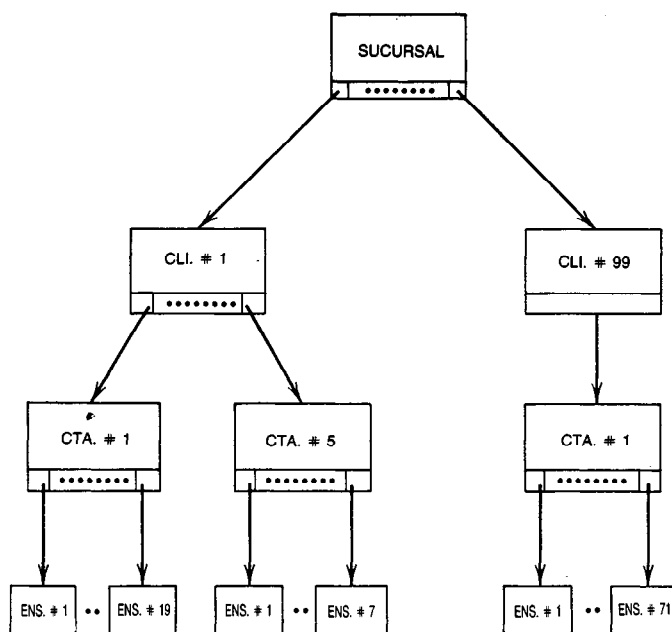


Fig. 57-7. Base de datos de un banco. CLI. = clientes; CTA. = cuenta; ENS. = ensayo.

tiene información acerca de una sucursal de banco. Los niveles inferiores contienen datos de clientes, cuentas de clientes y transacciones de cuentas, respectivamente. Cada componente posee o apunta a un número variable de componentes del siguiente nivel inferior de la jerarquía. Por ejemplo, cada componente de cliente apunta a las componentes de cuenta asociadas con ese cliente. Las características que definen este tipo de árbol son las siguientes:

1. Existe un punto de entrada o captación único llamado raíz.
2. Cada componente tiene cero o más componentes sucesoras inmediatas.
3. Existe una trayectoria de acceso o pista de apuntadores única de la raíz a una componente cualquiera.
4. No hay ciclos o retrocesos de la trayectoria de acceso en sí.

Por lo general se utiliza la terminología de los árboles genealógicos. Si una componente A apunta a una componente B, entonces A se llama padre de B y B se conoce como hijo de A. Las componentes con un padre común se llaman hermanos y juntos conforman una familia. Los ancestros y descendientes de las componentes se definen en la forma acostumbrada. Junto con sus descendientes, cada componente determina un subárbol del árbol original. Las componentes sin hijos se denominan puntos extremos. Si los hijos de cada componente se ordenan por rango, el árbol se denomina árbol ordenado<sup>8</sup>.

En la figura 57-8 se ilustra un árbol binario. El apuntador de una sola entrada se llama ROOT (raíz). Ca-

da componente tiene exactamente dos campos de apuntador, LPTR (apuntador izquierdo) y RPTR (apuntador derecho), donde uno u otro o ambos pueden ser nulos. Los descendientes que están a la izquierda de una componente forman el subárbol izquierdo de esa componente. Su subárbol derecho se define en forma análoga. Los árboles binarios son importantes en virtud de su uso como árboles de búsqueda y para representar árboles ordenados arbitrarios.

**Uso apropiado.** Los árboles resultan adecuados cuando un conjunto de datos es naturalmente jerárquico o, en un entorno dinámico, cuando las búsquedas son frecuentes. En el ejemplo del banco de la figura 57-7, la relación jerárquica entre las componentes es una característica intrínseca de los datos. El uso de una estructura de datos que modele esta organización tiende a simplificar la programación. Los algoritmos estándares para árboles pueden implementar de manera eficiente acciones jerárquicas normales con los datos, tales como la inserción de una nueva transacción bajo una cuenta. La organización enlazada de un árbol también da cabida a los requisitos de almacenamiento impredecibles originados por el número de componentes altamente variable en cada nivel del árbol<sup>19</sup>.

Un árbol binario es apropiado cuando un conjunto de datos es dinámico; es decir, cuando con frecuencia se insertan, suprimen y buscan componentes con base en alguna clave. Una clave es un valor de referencia que únicamente identifica una componente. Un ejemplo es el número del seguro social en un registro de personal, o el número de cuenta bancaria. Un árbol de búsqueda binario está definido por la propiedad de que la clave de cada componente es mayor que todas

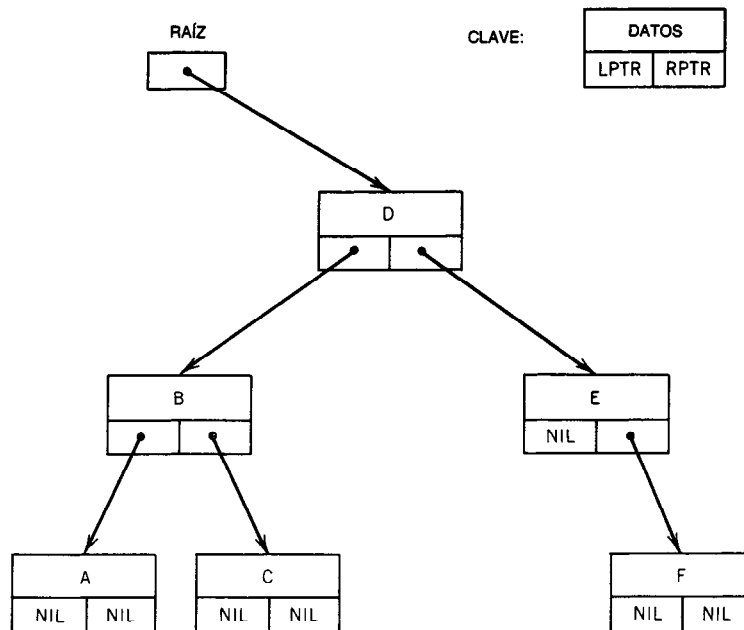


Fig. 57-8. Árbol binario. LPTR = apuntador izquierdo; RPTR = apuntador derecho.

las claves del subárbol izquierdo de esa componente y menor que todas las claves de su subárbol derecho. El árbol binario de la figura 57-8 es un ejemplo, suponiendo que el campo de datos es la clave. Esta disposición facilita la búsqueda de una componente en el árbol con una clave dada. Comenzando en la componente de la entrada, se compara la clave de búsqueda con la clave de una componente y se desplaza a la izquierda o derecha en el árbol, según si la clave de búsqueda es menor o mayor que la clave de la componente inspeccionada. La búsqueda termina con éxito cuando se halla una coincidencia, o con resultados negativos cuando se va más allá de un punto extremo. De aquí en adelante se presentan procedimientos formales para manipular árboles de búsqueda binarios. En las referencias 7, 20 y 21 pueden encontrarse mayores detalles.

**Representación.** Un árbol ordenado puede transformarse en un árbol binario por medio de un procedimiento llamado correspondencia natural<sup>8</sup>. Un árbol ordenado y su árbol binario correspondiente se ilustran en la figura 57-9. El primer paso del procedimiento consiste en entrelazar las componentes binarias de tal manera que la información estructural del árbol ordenado original esté implícita en el árbol binario correspondiente. Como ejemplo ilustrativo del proceso, supóngase que A es una componente del árbol ordenado y que A' es su componente binaria asociada (fig. 57-9). Sea B el primer hijo de A, y B' la componente binaria asociada con B. Sea C el siguiente hermano de A, y C' la componente binaria asociada con C. Los apuntadores de la componente binaria A' se dirigen de manera que el apuntador izquierdo de A'

[LPTR(A')] apunte a B' y RPTR(A') apunte a C'. Si se repite con todas las componentes, esto produce el árbol binario que corresponde al árbol ordenado. La importancia de este procedimiento es que permite limitar la atención a los árboles binarios.

En la figura 57-10 se muestra una representación del árbol binario de la figura 57-8 mediante el uso de arreglos. El patrón es similar al que se utiliza con una lista enlazada lineal, salvo por el hecho de que se utilizan dos campos de apuntador, LPTR y RPTR. La lista libre asociada (componentes 7 y 5) se implementa como una lista enlazada con sus apuntadores integrados en LPTR (los valores de RPTR correspondientes se pasan por alto).

Los árboles suelen representarse como estructuras de datos enlazadas, pero algunas veces se utilizan representaciones secuenciales más compactas. Estas son apropiadas si el almacenamiento es muy limitado o si el árbol se recorre secuencialmente con frecuencia, y mientras la forma del árbol sea estática. Por ejemplo, un árbol ordenado arbitrario puede representarse en forma compacta mediante el uso de dos bits por componente para almacenar información estructural<sup>2</sup>.

**Operaciones con árboles de búsqueda binarios.** Por lo general la clave es un campo del área de datos de una componente. Sin embargo, para simplificar, se supone que el área de datos misma es la clave. En los algoritmos se usa la representación del arreglo para los árboles de la figura 57-10.

**Búsqueda.** En la figura 57-11a se ilustra el pseudocódigo del procedimiento de búsqueda. En la salida, P es NIL (nulo) o apunta a donde se halló la clave.

**Inserción.** Antes de agregar un miembro al árbol de búsqueda, se tiene que buscar una componente con una clave duplicada. Si existe una, se rechaza el intento de inserción. De lo contrario, se inserta el nuevo miembro al final de la trayectoria de búsqueda fallida, como hijo de la última componente en esa trayectoria. La búsqueda que se requiere es una ligera modificación del procedimiento de búsqueda básico, y desplaza un par de apuntadores padre e hijo (Q y P) a través del árbol, de manera que a la salida P apunte a la clave duplicada o bien Q apunte al punto extremo donde falla la búsqueda. El pseudocódigo del procedimiento de inserción se ilustra en la figura 57-11b.

**Supresión.** El procedimiento de supresión es complicado. Las acciones que se requieren dependen considerablemente de circunstancias tales como la supresión de una componente que sea una raíz, la supresión de una componente con subárboles izquierdo o derecho vacíos, etc. El primer paso consiste en aplicar el procedimiento de búsqueda a partir de INSERT para localizar el candidato a la supresión (P) y su padre (Q). Si la búsqueda falla, no se requiere ninguna otra acción. De lo contrario, se hace que LINK(Q) se refiera a LPTR (Q), RPTR (Q) o ROOT, en la forma debida, cuando P sea el hijo izquierdo o derecho de Q o la componente de entrada del árbol. El pseudocódigo del procedimiento para suprimir la componente P se muestra en la figura 57-11c. La componente suprimida puede agregarse a una lista libre para volver a utilizarla.

**Recorrido.** El recorrido de un árbol es la visita de todas sus componentes. Los procedimientos de recorrido se utilizan para desplegar el contenido de un árbol en algún orden apropiado. También se requieren para buscar una componente cuando el parámetro de búsqueda no sea la clave según la cual el árbol está organizado. En ese caso el procedimiento de búsqueda binario resulta inaplicable y se requiere una exploración potencialmente de todas las componentes.

Las componentes de un árbol se pueden visitar en distintos órdenes. En el caso de árboles de búsqueda binarios, el recorrido "en orden" es el más apropiado porque permite visitar las componentes según el orden de la clave. En este tipo de recorrido, primero se despliega el subárbol izquierdo de una raíz, después la raíz y por último el subárbol derecho. Por ejemplo, en el caso de la figura 57-8, las componentes del subárbol con raíz en B desplegadas en orden son A,B,C. Las componentes del subárbol con raíz en E desplegadas en orden son E,F. El árbol completo, desplegado con la disposición en orden, es A,B,C,D,E,F. En la figura 57-12 se presenta el pseudocódigo de un procedimiento de recorrido en orden. Se emplea una pila, inicialmente vacía, para el almacenamiento temporal de los apuntadores. Knuth<sup>8</sup> describe otros órdenes de recorrido y más aplicaciones.

Los árboles también se pueden recorrer sin pilas, con ahorros de almacenamiento consecuentes, pero al precio de mayor complejidad algorítmica. Los árboles eslabonados son un dispositivo estándar para facilitar el recorrido cuando no se usan pilas<sup>9</sup>.

**Rendimiento.** La altura de un árbol de búsqueda determina su rendimiento. La altura se define como la longitud o el número de componentes de la trayectoria de búsqueda más larga. La altura de un árbol binario bien equilibrado con  $N$  componentes es del orden de  $\log_2(N)$ . El número de pasos que se siguen para buscar, insertar o suprimir elementos es determinado principalmente por la altura del árbol. En consecuencia, estos procedimientos requieren un promedio de  $\log_2(N)$  pasos en un árbol bien equilibrado<sup>7</sup>. El rendimiento se deteriora si el árbol crece de manera desequilibrada. Si ocurre esto, es posible reorganizar el árbol para equilibrarlo.

Alternativamente, se pueden utilizar árboles AVL<sup>2</sup> (en honor de sus creadores, Adelson-Velskii y Landis). Los árboles AVL utilizan procedimientos de inserción

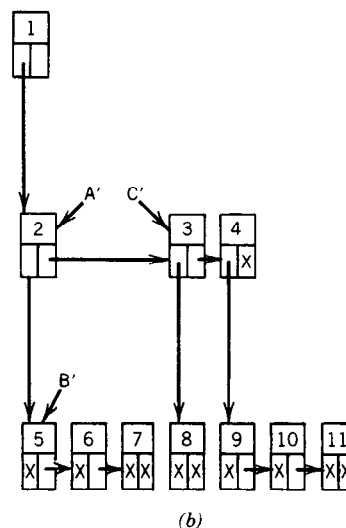
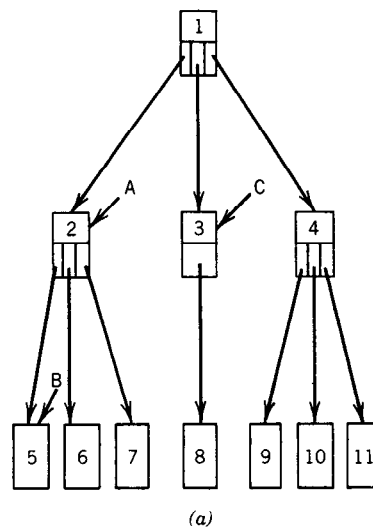


Fig. 57-9. Correspondencia natural: a) árbol ordenado; b) árbol binario correspondiente.



	DATA	LPTR	RPTR	
1	F	0	0	ROOT
2	D	3	4	2
3	B	6	8	
4	E	0	1	
5	—	0	—	
6	A	0	0	FREE
7	—	5	—	7
8	C	0	0	

Fig. 57-10. Representación con arreglos de un árbol binario.

y supresión que también efectúan operaciones de equilibrio en el árbol para conservar su altura próxima al mínimo.

Los árboles pueden organizarse asimismo para aprovechar información anterior acerca de frecuencias de búsqueda. Específicamente, dada una población completamente estable de componentes con frecuencias de búsqueda conocidas, se puede construir un árbol de búsqueda óptimo que minimice la longitud esperada de las trayectorias de búsqueda. La construcción requiere  $N^2$  pasos en el caso de una población de  $N$  componentes<sup>7</sup>.

Un árbol  $m$ -ario permite la existencia de  $m$  hijos debajo de cada componente del árbol. Un tipo de árbol  $m$ -ario, llamado árbol B, es útil para almacenar grandes índices estructurados en árboles en dispositivos de almacenamiento rotatorio como discos. Un árbol binario resulta inapropiado en un dispositivo de este tipo, porque las  $N$  componentes del árbol se pueden diseminar en muchas pistas. Si se utiliza un árbol binario, un promedio de  $\log_2(N)$  componentes yacen en la trayectoria de búsqueda. Estas componentes pueden situarse en diferentes pistas, y para el acceso a cada pista se requiere una operación extra de búsqueda en el disco. Por ejemplo, si  $N=16 \times 10^6$ , entonces  $\log_2(N)$  es aproximadamente 24. Por tanto, se requeriría un promedio de 24 búsquedas en el disco para obtener una componente. Para un árbol B con  $m$  hijos debajo de cada componente, la longitud de la trayectoria de búsqueda esperada es sólo de  $\log_m(N)$ . Si se utiliza un árbol B con  $m=24$ ,  $\log_{24}(N)$  es aproximadamente 5. De este modo sólo se requieren tres búsquedas para localizar una componente típica. Las operaciones de mantenimiento son más complicadas que para un árbol binario, pero esto se compensa con los ahorros que se logran en el tiempo de I/O debido al menor número de búsquedas<sup>7</sup>.

## 57.5 TABLAS DE ASIGNACIÓN

### 57.5.1 Definición

Una tabla de asignación consta de:

1. Un bloque de almacenamiento dividido en componentes del mismo tamaño.
2. Una función que sitúa las claves en las direcciones de componentes en el bloque.

Si  $K$  es una clave y  $H$  es una función de asignación, entonces  $H(K)$  da la dirección de una posición de tabla llamada dirección relativa de  $K$ . Una componente con esa clave se inserta en la posición y se encuentra ahí en el procedimiento de recuperación. Esto representa el mejor rendimiento posible de inserción y búsqueda que puede lograr una estructura de datos.

Por ejemplo, considérese una tabla que consta de espacio para 1 000 componentes, cada una con 100 bytes de longitud. Supóngase que las claves de las entradas por insertar en la tabla son identificadores de 10 caracteres. Defínase una función de asignación  $H$  de una clave  $K$  por medio de la fórmula:

$$H(K) = T + 100 \cdot V(K)$$

donde  $T$  es la dirección inicial de la tabla y  $V(K)$  es una función que asigna un valor numérico a  $K$ . Para que  $H(K)$  represente la dirección inicial de una componente de la tabla,  $V$  debe estar entre 0 y 999. Enseguida se presenta una elección ingenua de  $V$ , pero que ilustra algunos aspectos básicos. Supóngase que los caracteres se representan en forma interna en ocho bits, como en EBCDIC (*extended binary coded decimal interchange code*). Defínase  $V(K)$  como el valor decimal de, por ejemplo, el primer carácter de  $K$ . Si  $K$  es "SMITH, JOHN", entonces el primer carácter es S. Su representación en EBCDIC es 11000101, que tiene un valor decimal de 197. Por tanto  $H$  asociaría la clave con la posición 197 de la tabla. La entrada correspondiente se insertaría en esa posición.

En realidad el procedimiento de inserción no es tan simple. Surgen complicaciones debido a que diferentes claves no necesariamente se asignan a domicilios distintos. En realidad eso suele ser indeseable e inalcanzable. Es inalcanzable porque el número de claves suele ser enorme. Si distintas claves se situaran en posiciones diferentes, el área de almacenamiento tendría que ser inmensa. Es indeseable puesto que la mayor parte de las claves posibles *a priori* no ocurrirían en un problema dado, y de este modo nunca se utilizaría la mayor parte de esa inmensa tabla. En consecuencia, se espera que las funciones de asignación sitúen algunas claves en posiciones base relativas idénticas.

Cuando la clave de una nueva entrada se asigna a una posición ya ocupada, se dice que ocurre una colisión. En el ejemplo, esto le sucede a un identificador cualquiera con el mismo primer carácter. Una función de asignación bien diseñada deberá minimizar la posibilidad de que eso ocurra. La función de asignación del ejemplo tiene un desempeño deficiente en este sentido, en virtud de que sólo direcciona cerca de una cuarta parte de las direcciones disponibles en la tabla (256 de 1 000 localidades posibles). Esto obliga a las nuevas entradas a acomodarse en una pequeña porción del espacio disponible y de este modo aumenta la probabilidad de colisión.

Sin embargo, es de esperar que haya colisiones, y de este modo debe existir algún criterio para resolverlas cuando ocurran. Una solución simple consiste en colocar una entrada en colisión en la siguiente componente secuencial disponible. Una vez más, la función

de asignación del ejemplo funciona en forma deficiente con este criterio de colisión, ya que no sólo no direcciona la mayor parte de la tabla, sino que además todas las componentes direccionadas son contiguas (se ubican en las primeras 256 localidades). En el criterio de colisión de espacio siguiente disponible se utiliza precisamente esta misma porción de la tabla para manipular las colisiones. Esto incrementa la oportunidad de que haya colisiones subsiguientes.

Es claro que para utilizar una tabla de asignación deben considerarse con cuidado la elección de una función de asignación y el criterio de resolución de colisiones<sup>2,22</sup>.

### 57.5.2 Uso apropiado

Una tabla de asignación es la estructura de datos recomendada para problemas que requieren principalmente inserciones y recuperaciones conducidas por clave, pero pocas supresiones o recorridos ordenados por clave.

El rendimiento en cuanto a acceso suele ser excelente. Cuando se eligen la función de asignación y el procedimiento de manejo de colisiones correctos, son de esperar sólo una o dos colisiones por inserción o recuperación, aun cuando la tabla esté llena al 90%<sup>2</sup>. Una desventaja es que las tablas de asignación no

---

```
Set P := ROOT and FOUND := FALSE.
```

```
WHILE P ≠ 0 AND FOUND = FALSE DO:
```

```
  IF DATA(P) = SEARCHKEY
```

```
    set FOUND := TRUE, (* SALIDA. *)
```

```
  ELSE
```

```
    IF DATA(P) > SEARCHKEY
```

```
      set P := LPTR(P). (* MOVER A LA IZQUIERDA. *)
```

```
    ELSE
```

```
      set P := RPTR(P). (* MOVER A LA DERECHA. *)
```

```
  END DO.
```

(a)

```
Set P := ROOT, Q := 0, and FOUND := FALSE.
```

```
WHILE P ≠ 0 AND FOUND = FALSE DO:
```

(\* REALIZAR LA BÚSQUEDA? \*)

```
  IF DATA(P) = SEARCHKEY
```

(\* PROBAR SI HAY DUPLICIDAD. \*)

```
    set FOUND := TRUE,
```

```
  ELSE
```

(\* AVANZAR LOS APUNTADES. \*)

```
    set Q := P.
```

```
    IF DATA(P) > SEARCHKEY
```

```
      set P := LPTR(P).
```

```
    ELSE
```

```
      set FOUND := TRUE,
```

```
  END DO.
```

```
IF FOUND = FALSE
```

(\* REALIZAR LA INSERCIÓN. \*)

```
  Set NEW := FREE and FREE := LPTR(FREE).
```

(\* ASIGNAR ESPACIO. \*)

```
  Set DATA(NEW) := SEARCHKEY.
```

(\* INTRODUCIR DATOS. \*)

(b)

---

**Fig. 57-11.** Algoritmos de árboles binarios: a) búsqueda; b) inserción. Set = hacer; and = y.

```

Set LPTR(NEW) := 0 and RPTR(NEW) := 0.      (* FORMAR EL PUNTO EXTREMO. *)

IF Q = 0                                     (* ENLAZAR Q A P. *)
    set ROOT := NEW.

ELSE
    IF DATA(Q) < SEARCHKEY
        set RPTR(Q) := NEW.
    ELSE
        set LPTR(Q) := NEW.
    ELSE
        (* RECHAZAR LA INSERCIÓN *)
        COLOCAR EL MENSAJE DE CLAVE DUPLICADA
        (b) (cont.)

IF LPTR(P) = 0 OR RPTR(P) = 0               (* ¿ES P UN PUNTO EXTREMO? *)
    IF LPTR(P) = 0
        set LINK(Q) := RPTR(P),
    ELSE
        set LINK(Q) := LPTR(P).
    ELSE
        (* P NO ES UN PUNTO EXTREMO. *)
        Set C := LPTR(P).
        IF RPTR(C) = 0
            Set LINK(Q) := C,
            set RPTR(C) := RPTR(P),
        ELSE
            Set R' := C and R := RPTR(C).    (* DETERMINAR R, LA COMPONENTE MAS A
            WHILE RPTR(R) ≠ 0 DO:             LA DERECHA DEL SUBÁRBOL IZQUIERDO
                                                DE P. Y EL PADRE R' DE R. *)

            Set R' := R and R := RPTR(R).

        END DO.
        Set DATA(P) := DATA(R).           (* SUSTITUIR P POR R. *)
        Set RPTR(R') := LPTR(R).           (* SUPRIMIR R EN LUGAR DE P. *)
        (c)

```

Fig. 57-11 (cont.). Algoritmos de árboles binarios; b) inserción; c) supresión. Set = hacer; and = y.

pueden extenderse con facilidad. Si la tabla se llena y se requiere una tabla mayor, se deberán volver a asignar todas las entradas existentes.

Una aplicación típica de las tablas de asignación es como tablas de símbolos en ensambladores o compiladores<sup>23</sup>. Los ensambladores y compiladores deberán poder identificar los símbolos empleados en los pro-

gramas y almacenar esos símbolos y sus atributos para referencia posterior. Los símbolos pueden ser nombres de variables, símbolos mnemotécnicos de códigos de operación de un lenguaje ensamblador, etc. Entre los atributos se cuentan características de los símbolos tales como su tipo de datos (en relación con un nombre de variable) o su representación en código de máquina

---

```

Set P := ROOT and EXIT := FALSE.

REPEAT UNTIL EXIT = TRUE:

    WHILE P ≠ 0 DO:
        CARGAR P EN LA PILA.
        Set P := LPTR(P).
    END DO.

    Si pila = vacía
        (* ¿HAY ALGUNA COMPONENTE QUE NO SE HAYA VISITADO? *)
        set EXIT := TRUE,
    ELSE
        Descargar la cima de la pila en P
        Visitar DATA(P).
        Set P := RPTR(P).
    END REPEAT

```

---

Fig. 57-12. Recorrido de un árbol "en orden". Set = hacer; and = y.

(para un código de operación ensamblador). Un conjunto de entradas de este tipo, donde cada una consta de un símbolo y sus atributos, constituye una tabla de símbolos. El símbolo mismo es la clave de la entrada. Suelen predominar las inserciones y recuperaciones, y de este modo resulta apropiada una tabla de asignación. Por otra parte, las tablas de símbolos para nombres de variables en lenguajes estructurados en bloques, como Pascal y Algol, están sujetas asimismo a supresiones. En ese caso un árbol de búsqueda binario equilibrado puede resultar la estructura de datos más adecuada. Si existe un requisito de tiempo real, se puede optar por un árbol AVL debido al mal comportamiento en el peor de los casos de una tabla de asignación cuando se llena hasta el tope (más del 95%) y debido a la dificultad que existe para extender la tabla.

### 57.5.3 Funciones

Una función de asignación adecuada deberá ser fácil de calcular y diseminar en forma aleatoria las claves esperadas en la tabla, minimizando el riesgo de colisión. Si la clave es lo suficientemente pequeña para ser operada por las instrucciones aritméticas de la máquina, a menudo el método de la división resulta apropiado:

$$H(K) = T + S*(K \text{ MOD } P)$$

donde  $T$  es la dirección inicial de la tabla;  $S$  es el tamaño de la componente, en bytes;  $K$  es la clave; y  $P$  es un número primo grande<sup>22</sup>. La tabla correspondiente, como es natural, deberá tener espacio para alojar  $P$  componentes. La teoría y la experiencia indican que la elección de  $P$  como un número primo grande minimiza

el riesgo de colisión. Si la clave es demasiado grande para ser manipulada directamente por instrucciones aritméticas, entonces se puede aplicar el desplazamiento múltiple<sup>22</sup>. Esto divide la clave de  $N$  bits en subgrupos más pequeños de  $M$  bits y después suma o aplica la operación OR excluyente a estos grupos de  $M$  bits en un resultado final de  $M$  bits. Esto proporciona efectivamente un índice en una tabla con  $2^M$  componentes. El desplazamiento múltiple se utiliza algunas veces meramente para precondicionar o reducir el tamaño de la clave hasta el punto en que se pueda aplicar una función de asignación aritmética, tal como el método de la división. Si se sabe que las claves esperadas exhiben algún patrón regular, se puede aplicar el análisis de cifras<sup>24</sup>. Este método selecciona caracteres de posiciones fijas en una clave, de manera que la subclave resultante sea lo suficientemente pequeña para utilizarse como índice de la tabla y lo suficientemente aleatoria para minimizar las colisiones.

### 57.5.4 Resolución de colisiones

El encadenamiento y direccionamiento abierto son las técnicas básicas para el manejo de colisiones. También se aplica una técnica llamada de almacenamiento en canastillas para tablas de asignación en dispositivos de almacenamiento externos, pero no se considera aquí<sup>2</sup>.

**Encadenamiento.** En esta técnica la entrada en colisión se inserta en una lista llamada de desbordamiento. Se conserva una lista de desbordamiento aparte por cada posición de la tabla. El espacio para las componentes de las listas de desbordamiento se asigna a partir de una lista libre, que está separada de la tabla de asignación principal.

**Direccionamiento abierto.** En esta técnica el procedimiento de colisión busca en forma sistemática una posición libre en la tabla de asignación. Si se encuentra una, ahí se hace la inserción. De lo contrario, se rechaza la solicitud de inserción. La secuencia de búsqueda se puede generar de maneras diferentes<sup>2</sup>.

Existe una componenda entre el tiempo que lleva calcular una función de asignación y el grado de riesgo de colisión. En una tabla almacenada internamente puede resultar más costoso determinar una función de asignación compleja con propiedades de evasión de colisiones superiores que meramente manejar las colisiones cuando éstas ocurran. En el caso de una tabla almacenada externamente, el costo de la I/O extra durante la resolución de colisiones bien puede superar el costo de calcular una función de asignación.

### 57.5.5 Procedimientos de acceso

Los procedimientos para insertar, buscar y suprimir una entrada se exponen a continuación.

**Inserción.** La inserción requiere asignación y quizá resolución de colisiones. La clave de la entrada por insertar se asigna a una dirección de destino. Si se dispone de la componente de destino, ahí se inserta la entrada. De lo contrario, se utiliza un procedimiento de resolución de colisiones. Si el procedimiento logra identificar una posición disponible alternativa, ahí se hace la inserción. De lo contrario, la tabla se identifica como llena y se rechaza la solicitud de inserción.

**Búsqueda.** En este procedimiento se asigna primero la clave de búsqueda. Si se encuentra una clave coincidente en esa posición base relativa, se detiene la búsqueda, con buenos resultados. Si la posición base relativa está ocupada, se detiene la búsqueda, sin cumplir su objetivo, debido a que la clave no está en la tabla. De lo contrario, el procedimiento de búsqueda da inicio a la exploración de la cadena de resolución de colisiones que comienza en la posición base relativa. Este barrido exploratorio se suspende cuando se encuentra una clave que coincide o bien si la cadena termina antes de que ocurra una coincidencia. Con el encadenamiento, la cadena de resolución de colisiones es una lista de desbordamiento, y el barrido termina al final de esta lista o antes de que se encuentre una coincidencia. Con el direccionamiento abierto, el barrido se suspende si se encuentra una posición vacía (que indica el final de la cadena), si la cadena regresa a la posición base relativa (lo que indica que la entrada no está en la tabla) o si se encuentra una coincidencia.

**Supresión.** La supresión requiere cierto cuidado, ya que una posición de la cual se suprima una entrada puede ubicarse en una secuencia de resolución de colisiones de direccionamiento abierto. Si esta posición no se indicara, actuaría incorrectamente como una posición de suspensión en una búsqueda de direcciones abierta. Aunque se distinguen de las posiciones libres para la búsqueda, las posiciones suprimidas siguen estando disponibles para hacer inserciones<sup>2</sup>.

### REFERENCIAS BIBLIOGRÁFICAS

1. M. C. Harrison, *Data Structures and Programming*, Scott, Foresman, Glenview, IL.
2. T. A. Standish, *Data Structure Techniques*, Addison-Wesley, Reading, MA.
3. J. L. Pfaltz, *Computer Data Structures*, McGraw-Hill, Nueva York.
4. C. C. Gotlieb y L. R. Gotlieb, *Data Types and Structures*, Prentice-Hall, Englewood Cliffs, NJ.
5. M. V. Wilkes, "Lists and Why They Are Useful", *Comp. J.*, pp. 278-281.
6. W. L. Honig y C. R. Carlson, "Toward an Understanding of (Actual) Data Structures", *Comp. J.*, pp. 98-104.
7. N. Wirth, *Algorithms + Data Structures = Programs*, Prentice-Hall, Englewood Cliffs, NJ.
8. D. Knuth, *The Art of Computer Programming*, Vol. 1, *Fundamental Algorithms*, Addison-Wesley, Reading, MA.
9. E. Horowitz y S. Sahni, *Fundamentals of Data Structures*, Computer Science, Potomac, MD.
10. R. Williams, "A Survey of Data Structures for Computer Graphics Systems", *Comp. Surv.*, pp. 1-23.
11. W. Newman y R. Sproull, *Principles of Interactive Graphics*, McGraw-Hill, Nueva York.
12. J. H. Sexton, "An Introduction to Data Structures with Some Emphasis on Graphics", *Comp. J.*, pp. 444-447.
13. T. G. Lewis y M. Z. Smith, *Applying Data Structures*, Houghton Mifflin, Boston, MA.
14. J. P. Tremblay y P. G. Sorenson, *An Introduction to Data Structures with Applications*, McGraw-Hill, Nueva York.
15. K. J. Thurber y P. C. Patton, *Data Structures and Computer Architecture*, Lexington Books, Lexington, MA.
16. A. T. Bertziss, *Data Structures: Theory and Practice*, 2.<sup>a</sup> ed., Academic, Nueva York.
17. W. M. McKeeman, "Stack Computers", en H. S. Stone, ed., *Introduction to Computer Architecture*, 2.<sup>a</sup> ed., Science Research Associates, Chicago.
18. K. B. Magleby, "Introduction to Minicomputers", en H. S. Stone, ed., *Introduction to Computer Architecture*, 2.<sup>a</sup> ed., Science Research Associates, Chicago.
19. G. Wiederhold, *Database Design*, McGraw-Hill, Nueva York.
20. E. M. Reingold, J. Nievergelt y N. Deo, *Combinatorial Algorithms: Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ.
21. J. Nievergelt, "Binary Search Trees and File Organization", *Comp. Surv.*, pp. 195-207.
22. W. D. Maurer y T. G. Lewis, "Hash Table Methods", *Comp. Surv.*, pp. 5-20.
23. J. Donovan, *Systems Programming*, McGraw-Hill, Nueva York.
24. C. D. Knott, "Hashing Functions", *Comp. J.*, pp. 265-278.